

UNCLASSIFIED

AD NUMBER
AD405483
NEW LIMITATION CHANGE
TO Approved for public release, distribution unlimited
FROM Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; FEB 1963. Other requests shall be referred to Office of Naval Research, Arlington, VA.
AUTHORITY
ONR ltr, 27 Feb 1974

THIS PAGE IS UNCLASSIFIED

UNCLASSIFIED

AD 405 483

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA



UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

NO.OTS

405 483

405483

FILE COPY

REFERENCE COPY

1

**Best
Available
Copy**

General

REFERENCE COPY

US Govt Printing Office
Price \$4.00 per
Telecon w/ONR 3 Tmbs
TISIA - 1 WMO

(6) PERSPECTIVES

REFERENCE COPY

IN

MATERIALS RESEARCH

Edited by

L. HIMMEL

*Department of Mineral
Technology
University of California,
Berkeley*

*(Formerly with Office of Naval
Research)*

J. J. HARWOOD

*Manager, Metallurgy
Department
Scientific Laboratory, Ford
Motor Co.*

*(Formerly with Office of Naval
Research)*

W. J. HARRIS, Jr.

*Assistant to the Vice President
Battelle Memorial Institute*

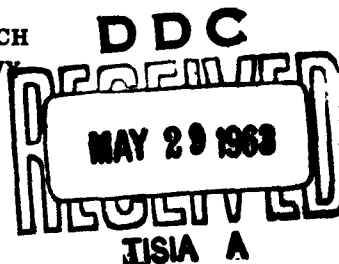
*(Formerly with the National Academy of Sciences—National
Research Council)*

Organized and Written by

**Advisory Committee on Perspectives in Materials Research
Division of Engineering and Industrial Research
National Academy of Sciences—National Research Council**

Sponsored by

**OFFICE OF NAVAL RESEARCH
DEPARTMENT OF THE NAVY
WASHINGTON, D.C.**



**ADVISORY COMMITTEE ON PERSPECTIVES
IN MATERIALS RESEARCH**

National Academy of Sciences—National Research Council

Prof. FREDERICK SEITZ, Chairman
*University of Illinois
Urbana, Ill.*

Members

Prof. HARVEY BROOKS
*Harvard University
Cambridge, Mass.*

Prof. BRUCE CHALMERS
*Harvard University
Cambridge, Mass.*

Dr. MORRIS COHEN
*Massachusetts Institute of
Technology
Cambridge, Mass.*

Prof. POL DUWEZ
*California Institute of
Technology
Pasadena, Calif.*

Dr. J. E. GOLDMAN
*Ford Motor Co.
Dearborn, Mich.*

Prof. ROBERT GOMER
*University of Chicago
Chicago, Ill.*

Dr. CONYERS HERRING
*Bell Telephone Laboratories
Summit, N. J.*

Prof. DAVID LAZARUS
*University of Illinois
Urbana, Ill.*

Prof. ROBERT J. MAURER
*University of Illinois
Urbana, Ill.*

Prof. E. R. PARKER
*University of California
Berkeley, Calif.*

Dr. DAVID TURNBULL
*General Electric Company
Schenectady, N. Y.*

Dr. GEORGE VINEYARD
*Brookhaven National
Laboratory
Upton, Long Island, N. Y.*

CONTENTS

	Page
Preface.....	vii
Introduction.....	x
Part I. THE SCIENCE OF MATERIALS.....	1
1. Relation of Science to Technology.....	2
2. Traditional Fields versus the Science of Materials.....	4
3. Growth of the Science in the Last Half-Century.....	6
4. Fields and Distribution of Research Activity.....	8
5. Interrelations of Different Fields.....	14
6. The Problem of Perspective.....	16
Part II. COHESIVE PROPERTIES OF SOLIDS.....	19
1. Introduction.....	21
2. Rigorous Theory of Cohesion for Pure Elements and Compounds.....	22
3. Elastic Constants and Interatomic Forces.....	28
4. Quantum Basis of Crystal Chemistry.....	33
5. Theory of the Cohesive Properties of Alloys.....	40
6. Methods of Measurement of Cohesive Properties.....	45
7. References.....	51
8. Appendix: Cohesion of Metallic Alloys—A Short Theoretical Survey.....	52
Part III. MAGNETISM AND MAGNETIC MATERIALS.....	75
1. Introduction.....	76
2. Paramagnetism and Paramagnetic Materials.....	79
3. Cooperative Magnetic Interactions.....	86
4. Static Properties of Ferromagnetic and Ferrimagnetic Materials.....	107
5. Dynamic Properties of Ferromagnetic and Ferrimagnetic Materials.....	129
6. Diamagnetism and Superconductivity.....	142
7. Bibliography and References.....	147
Part IV. ELECTRICAL, OPTICAL, AND THERMAL PROPERTIES OF SOLIDS.....	155
1. Electron and Phonon Processes in Solids.....	158
2. Dielectric Breakdown in Solids.....	171
3. Ferroelectricity.....	179
4. Thermoelectricity.....	191
5. Thermal Conductivity.....	200
6. Electronic Optical Absorption.....	222
7. Luminescence.....	237
Part V. DIFFUSION AND MASS TRANSPORT IN SOLIDS.....	258
1. Introduction.....	254
2. Experimental Techniques.....	259
3. Atomic Models.....	267
4. Phenomenological and Semi-Empirical Models.....	281

5. Problem Areas	287
6. References	308
Part VI. PHASE TRANSFORMATIONS IN THE SOLID STATE	309
1. Introduction	312
2. The Chemical Driving Force	315
3. Pre-Transformation Phenomena	323
4. Crystallography	326
5. Order-Disorder and Clustering Reactions	336
6. Nucleation	343
7. The Growth Process	349
8. General Kinetics of Solid-State Transformations	370
9. Summary	376
10. Bibliography	380
Part VII. GROWTH, STRUCTURE, AND MORPHOLOGY OF CRYSTALS	383
1. Introduction and Acknowledgments	385
2. Historical Survey	387
3. Crystal Chemistry	390
4. Theory of Growth and Morphology	397
5. Control of Purity and Structure	407
6. Crystallization of Polymers	430
7. Formation and Structure of Thin Films	435
8. Summary of Conclusions and Current Important Problem Areas	444
9. References	450
Part VIII. MECHANICAL BEHAVIOR OF CRYSTALLINE SOLIDS	455
1. Introduction	456
2. Dislocation Theory—Its Current State of Development and Limitations	458
3. Experimental Observations on Flow and Fracture	469
4. Anelasticity and Internal Friction	484
5. Conclusions and Suggestions for Future Research	492
6. References	494
Part IX. SURFACE PHENOMENA—THE NATURE AND PROPERTIES OF SOLID SURFACES AND INTERFACES	498
1. Introduction	501
2. Thermodynamics of Surfaces	511
3. Surface Structure	518
4. The Role of Imperfections in Surface Phenomena	523
5. Structure and Properties of Grain Boundaries	539
6. Friction, Adhesion, and Lubrication	551
7. Electronic Surface Properties of Metals and Semiconductors	566
8. Interactions of Surfaces with Gases	581
9. Liquid Electrode Reactions and Corrosion	613
10. A Summary of Unsolved Problems	626
Part X. STRUCTURE AND PROPERTIES OF LIQUIDS	628
Part XI. EFFECTS OF RADIATION ON MATERIALS	629
1. Introduction	631
2. Assessment of the Present Status of the Field	634
3. Need and Prospects	667
4. Sputtering	699
5. References	707

CONTENTS

Part XII. TECHNIQUES AND INSTRUMENTATION	718
1. Introduction and Acknowledgments.....	718
2. Microtechniques and General.....	716
3. High Pressure, High Temperature.....	730
4. High Temperature.....	738
5. High Intensity Stress Waves.....	744
6. High Magnetic Fields.....	751
7. High Frequency Ultrasonics.....	758
8. Analytical Techniques.....	766

PREFACE

One of the areas of science which has received particular attention during the past 15 years has been research on the solid state. The advances in solid state physics, metallurgy, and chemistry have been of such magnitude that one is tempted to state that our understanding of the properties and behavior of materials has reached a high degree of sophistication—so much so, that the utilization of basic science in the improvement of materials will hopefully displace the successful, but relatively inefficient, empiricism of the past. Out of this has grown a multibased research discipline—materials science—an integration of the concepts of physics, chemistry, metallurgy, and mechanics which considers the properties of materials from atomistic, electronic, and structural relationships, rather than from arbitrary groupings of classes of materials.

One penalty of the outpouring of published material covering the accelerated research activities during the past years in almost every aspect of the science of materials has been the almost necessary self-defensive retreat of individual researchers toward specialization. Maintaining a proper awareness of the state of affairs in a broad field of science—even one as limited as the solid state or yet, more specifically, metallurgy—has become tremendously difficult, if not impossible. Moreover, the results of the scientific research conducted in the last two decades—an unparalleled period of intensive dedication to science and research—demonstrate that the boundaries to our knowledge are not readily circumscribed. It seems instead that the entropy of our understanding becomes only more delineated by the order of our increased scientific progress and knowledge. Several excellent review publications, symposia proceedings, etc., serve to reduce the plethora of scientific publications to a manageable state, but such reviews concentrate on summarizing, annotating, and referencing the state of available knowledge. Too infrequently is there any effort to provide a critical perspective in the appraisal of research in a field, although it is just as important to characterize the limitations and barriers to our understanding as to delineate the breadth of our knowledge.

During the past few years, the Office of Naval Research, through

its Metallurgy Branch, became increasingly concerned about this situation and the degree of specialization reflected in many of the Government-supported research programs, which comprise the bulk of the current national research activity. In view of the great expansion of the research effort in materials sciences and the difficulties for many to maintain a broad insight and perspective in the fields of endeavor encompassed by the "Science of Materials," it seemed important to question whether a proper share of the current research effort was being devoted to truly significant problems—whose solutions would enable a "quantum jump" in understanding the properties and behavior of materials. Or, in the words of C. P. Snow, "Was there an awareness on the part of many individual researchers, particularly the younger graduates, as to how big was the beach and how many pebbles were there, how many pebbles had been picked up, and how many were worth picking up?"

It appeared timely to consider the preparation of a comprehensive statement concerning the state of knowledge and the nature of as yet unresolved important problems in the several fields that comprise the science of materials. An informal group of leading scientists endorsed this concept and the National Academy of Sciences agreed on the need for a detailed study of the science of materials. The Division of Engineering and Industrial Research, National Academy of Sciences—National Research Council, was assigned responsibility for administering an ONR contract given to the NAS to provide funds for this study.

An Advisory Committee on Perspectives in Materials Research was appointed by the NAS-NRC, and this committee then designated the specific fields germane to the science of materials that should be included in the study. On the advice of the committee, panels of authors for the various specified fields were selected by the NAS. The studies were drafted and reviewed by these panels of authors during a period of about a year. Completed manuscripts were reviewed by members of the Advisory Committee and by other individuals expert in the field. Each panel exercised final editorial control over the contents of its own presentation. The material for this book was completed at various times during 1960.

The studies prepared by the various panels constitute the technical content of this volume. It was clear to the Advisory Committee in its original deliberations that not all fields of science pertinent to materials could feasibly be analyzed in the initial study. Many broad fields, such as organic chemistry and non-crystalline materials, were purposely omitted in order to provide emphasis on a rather smaller selection of topics.

No attempt was made to give equivalent attention to each of the fields under study. Some topics could be treated adequately only by the preparation of manuscripts several times longer than those included here. Nevertheless it does appear that the several parts of this volume taken individually and collectively represent a significant contribution to the understanding of the science of materials.

This book, then, is an attempt to provide perspective to the mushrooming field of materials sciences and to set forth for a number of topics the degree of understanding which has been achieved and the important problem areas which still plague us. If we are not in sight of the end, could we at least indicate the end toward which we might sight? It is our hope that such a perspective will be of particular assistance to the fresh new crop of research talent entering into this dynamic field of activity.

In no way should this book be considered as a rigid blueprint for the selection of new projects which might receive support from government contracting agencies. It is very likely that many ideas and concepts which have been omitted may become far more important and significant than those discussed. Nevertheless, this appraisal of recent achievements and critical unresolved problems in perspective should help guide scientists in independent thinking and new approaches—which are essential to enlightened and progressive advances.

The editors thank the authors and reviewers for their contributions and efforts. These men participated willingly and unselfishly, and without their labors this publication would not have been possible.

The gentlemen of the Advisory Committee also served as chairmen of the study panels or as authors of material in this book. Their assistance and guidance are much appreciated. Gratitude is extended to Dr. Shirleigh Silverman, Director of the Naval Research Group and to Dr. Immanuel Estermann, formerly Director, Material Sciences Division (currently, Chief Scientist, London Branch Office) of the Office of Naval Research for their continuous encouragement and support for this study program. A special note of thanks goes to Mr. I. Rudin, Technical Information Division, Naval Research Laboratory, for his never-failing assistance in the editorial and publication chores.

WASHINGTON, D. C.
JANUARY 1961.

J. J. HARWOOD,
L. HIMMEL,
W. J. HARRIS, Jr.,
Editors.



INTRODUCTION

FREDERICK SEITZ
University of Illinois

Until the development of the science of chemistry, the fundamental knowledge of materials was so intimately associated with the technology of preparation that fields such as those dealing with metals, ceramics, glasses, pigments, and fabrics could be regarded as self-contained. All aspects of each subject were closely interwoven. It is true that imaginative men of wide vision transferred knowledge from one field to another, but the typical member of a profession or craft could confine his attention to his own field from apprenticeship to masterhood. Innovations were inclined to be sporadic and their use restricted to local groups or schools as part of their specialized stock-in-trade. Under these conditions which were far from devoid of inspiration or genius, many of the most useful materials of the present day were discovered and fabricated.

The evolution of chemistry during the last century, the discovery of the elements, and the development of quantitative analytical techniques brought an end to the isolation of these fields. It became possible to subject materials to a new and powerful scrutiny from a unifying viewpoint. Thus, by 1900 the flow of knowledge from the field of chemistry into the areas devoted to materials represented by far the greatest source of innovation. This flow occurred both directly through the literature of chemistry and by explicit borrowing of the methods of chemical analysis.

By 1930 the good educational and research programs dealing with materials technology were in substantial part, although by no means entirely, aspects of applied chemistry. In each field there remained strong, essentially irreplaceable, components of the areas of older wisdom. The amount varied depending upon the degree to which chemical research could be of value and the extent to which the drive for technological progress would justify the expense of research. The year 1930 is not emphasized here because of any belief that the benefits which could be expected from applications of chemistry had leveled off at that time, for the contrary is true. Indeed, the vast development of polymer chemistry and

the applications of studies of rate processes in materials lay ahead. Rather, by 1930 any individual devoted professionally to the development of materials was at most an inspired technician if he did not have a thorough appreciation of the aspects of his field derived from chemistry. Moreover, many chemists had taken positions of leadership in developing and exploiting materials. Interestingly enough, this reliance upon chemistry is almost taken for granted at the present day even though it once was revolutionary. While the professional chemist does not have exclusive control of the development of materials, he is a thoroughly familiar and respected member of any team which is concerned with such development.

About 1930 another area of science began to have an impact on the technology of materials which ultimately proved to be comparable to that of chemistry. This area was that of atomic or molecular physics, particularly the broad region now bearing the designation of solid state physics. Actually, some of the products of the physicist, but rarely the man, had been welcomed into the areas devoted to the technology of materials. These products included, most notably, items of test equipment and physical standards, optical microscopes, spectrosopes, and the techniques of X-ray diffraction. However, the great devotion of the physicist to the laws of atomic structure had very little effect, except indirectly through chemistry.

By 1930 the fundamental laws of quantum mechanics had been formulated and were available for exploitation. Crystalline solids proved to be particularly amenable to systematic study through the use of the new mathematical and experimental tools. In fact, the study of such solids became one of the major areas of the field of physics. It should be emphasized that solid state physics actually was not a new field at that time for its roots can be traced very far back into the history of science. However, the field had now obtained new vitality and perspective because the underlying laws governing atomic behavior became available as a guide. The development which followed was far too broad and profound to be ignored by those interested in the development of materials. It included substantial progress in understanding topics such as the source of the cohesive energy of solids, the nature of electrical conductivity, and the origin of magnetism. In addition, the work provided substantial insight into the role that imperfections in solids play in determining their properties. By 1955 the key concepts of solid state physics were sufficiently highly developed that they were ready to become an important, indeed completely in-

dispensable, part of the background science from which the future technology would develop. Today no organization which presumes to play a vital role in the development of materials can afford to treat this aspect of the field as if it were only of peripheral interest.

What of the future? It is interesting to note that the marriage of chemistry and the traditional fields of materials technology, although complete, proceeded in a way which permitted those fields to retain their identity. The vocabulary in common use was altered substantially, and a fraction of the most competent practitioners started their careers as professional chemists. Nevertheless, fields such as metallurgy, ceramics, and glass technology are still well recognized. There is no reason to expect a serious change in the future as the field of physics exerts its full influence upon the development of materials, provided the minds of the leaders in the traditional fields prove sufficiently flexible to absorb the advances.

There is, however, one major way in which a change in organizational thinking may be necessary in the future. The number of individuals being trained in the fields of chemistry and physics with an explicit interest in materials is now sufficiently great to permit them to retain much of their professional association with their parent fields of science. For example, the physical or inorganic chemist, concerned with the study of materials, and the solid state physicist will continue to find the activities of their professional societies of interest to them in the future. This circumstance, coupled with the fact that the level of sophistication of science needed to make substantial progress will henceforth increase steadily, implies there will be need for very close cooperation between chemistry, physics, and the traditional fields of materials technology. In fact, one can scarcely imagine effective progress without strong interdisciplinary ties between fields, in a way which leads to the sharing of basic knowledge and techniques.

The present volume deals with these perspectives of materials research of the present and future provided by recent developments of basic science. It may be observed that the contributors represent a variety of disciplines as well as the interests of centers in government, industry, and the universities. Further emphasis on the interdisciplinary nature of the materials science is offered by the number of countries represented by the authors.

Part I

THE SCIENCE OF MATERIALS

Dr. CONYERS HERRING
Bell Telephone Laboratories
Murray Hill, N. J.

CONTENTS

	Page
1. RELATION OF SCIENCE TO TECHNOLOGY.....	2
2. TRADITIONAL FIELDS VERSUS THE SCIENCE OF MATERIALS	4
3. GROWTH OF THE SCIENCE IN THE LAST HALF-CENTURY	6
4. FIELDS AND DISTRIBUTION OF RESEARCH ACTIVITY.....	8
5. INTERRELATIONS OF DIFFERENT FIELDS.....	14
6. THE PROBLEM OF PERSPECTIVE.....	16

THE SCIENCE OF MATERIALS

This is a book on the science of materials. It differs from most other books in this area, in being designed not so much to expound facts and principles in detail as to present a panoramic view of the state of the science today, its recent progress, its deficiencies, and its possible future development.

These statements are simple to make, and they are accurate. But to those who, like most readers of this book, have concerned themselves with one or more of the fantastically diverse aspects of materials, many questions are sure to occur. Why talk of a "science of materials"? Where does it begin and end? Isn't it a rather big job to give a panoramic view of it? Why essay something so presumptuous, when there are plenty of good books and reviews? The task of this opening chapter will be to answer questions of this type, and to chart the gross pattern of which the more detailed perspectives of the later chapters are components.

There are a few things which can be said at once about the aims and limitations of the book. It covers only pure science, and not technological applications. But a significant part of the motivation for undertaking it arose from a feeling that the long-range progress of materials technology is coming to depend more and more on applied scientists with a broad perspective in pure science. This relation of science to technology, important not only for the technologist but also for the scientist, deserves further discussion, and will be the subject of the first of the sections to follow.

An effort has been made to map out areas of light and darkness with as little regard as possible for the necessary but often rather artificial barriers which separate physics, chemistry, metallurgy, etc., from one another. However, to keep the scale of the book manageable, some otherwise appropriate areas have had to be excluded. The principal areas omitted have been glasses and polymers; except for a chapter on liquids, a subject relevant to the understanding and preparation of solids, the entire book is limited to crystalline materials. Not quite all the properties of the latter are covered, either, and one of the tasks of this chapter will be to compare the coverage of the later chapters with the world-wide pattern of research effort.

1. RELATION OF SCIENCE TO TECHNOLOGY

We have mentioned the distinction between pure science and technology. The boundary between the two is certainly very nebulous. For example, research papers show a continuous spectrum of content, from the extreme abstractions of subatomic theory to very specific techniques for the manufacture of particular gadgets. For our purpose one may define pure science as the effort to understand an ever-widening variety of phenomena in terms of ever simpler and more tightly knit principles. Thus, one may say that pure science proceeds by collecting the specific and distilling out the general, while technology applies the general knowledge to specific situations, or, in the absence of general knowledge, solves specific practical problems by trial and error, with only a limited generalization. Clearly, though, these definitions apply only to the large-scale trends and aims of a research effort. In its day-to-day progress the path of pure science must take many ups and downs between the particular and the general, while technology often turns up generalizations that overlap the domain of science.

Granted, then, that we can at least roughly distinguish pure science from technology, let us examine the relation of these two

areas to each other. The statement that technology is based on science has been repeated so often that we have come to accept it without thinking just what it means. Actually, as every worker in the technology of materials knows, the scientific roots of a technology have often been very remote indeed from current technological progress. Consider, for example, the commercial development of copper-oxide and selenium rectifiers in the 1920's. This was an outgrowth of pure science in a remote sense: nineteenth-century science fathered both the electrical industry with its need for rectifiers and the original discovery of rectification. But the actual progress of rectifier development in the 1920's was made by trial and error, with little or no aid from current work in the pure sciences of physics and chemistry. Real progress in the scientific understanding of rectifier materials did not come until the 1930's. In countless fields of metallurgy, ceramics, etc., the story has been the same. It is not, therefore, superfluous to ask how relevant a book on current trends in pure science is likely to be to the technology of materials.

It is, of course, not hard to find examples of a different kind, where the development of commercial materials has been intimately dependent on current scientific understanding. The materials for modern silicon and germanium rectifiers and transistors are prepared by techniques suggested by recent advances in physics and chemistry, and their purity is monitored by methods worked out deductively from prior scientific knowledge. In another field, ceramics, still dominated by trial-and-error methods, one group of workers has recently used our newly acquired understanding of the atomic nature of the sintering process as the basis for a systematic program of development which has perfected transparent aluminum oxide ceramics. Examples of this sort are becoming more and more frequent, although there are still many areas of materials technology where one does better to rely on ad hoc experimentation and the tradition of particular crafts, rather than to try to deduce appropriate procedures from scientific knowledge. The useful properties of materials are often so extraordinarily sensitive to many factors—trace impurities, rate of cooling, surface conditions, etc.—that pure science has not yet learned how to take them all into account. In such cases technology must go it alone, and often succeeds in advancing ahead of science. But the trend of the times is the other way. As we shall see in some of the sections to follow, the pure science of materials has given us, in the last two decades or so, a breathtaking acceleration of our understanding. As the future unfolds,

we can expect science to provide substantial assistance to technology in more and more fields, and to lag behind it in fewer and fewer. In this expectation, and in the new technologies which the proliferation of science is suggesting, lies the interest of this book for those who devote themselves to the development of practical materials.

While it is true that science is catching up with technology, the relation of these two areas is not that of competitors in a race. When science lags behind technology the two may become separated from each other, each progressing independently. But when science has caught up, it not only provides the foundation for technology, but receives in return many powerful stimuli and influences. These are very diverse: there is the purely economic effect of increased support of certain areas of basic research, exemplified by the impact of the transistor on semiconductor research; there is the widening of research opportunities by application of new technologies to instrumentation, as has occurred with microwaves; finally, there are the posing of previously unrecognized problems, and the possibility of applying to problems of pure science some of the ingenious ideas developed by technologists for their own purposes.

2. TRADITIONAL FIELDS VERSUS THE SCIENCE OF MATERIALS

Materials are things which can exist in indefinite amount. They have what are called "bulk properties," properties which could in principle be measured anywhere within an infinite sea of the material in question. In other words, they are properties which bodies of this material possess independently of their size and shape—at least within wide limits. Similarly, one can define surface properties of a material as properties possessed by the interface between this material and empty space or some other material, which are practically independent of the dimensions of this interface. Certain properties may fall within or without one of these definitions, according to the scale of sizes one has in mind; for example, on a gross scale, many properties of a polycrystalline solid are bulk properties, while a fine-scale approach would show these properties to be dependent on the sizes and shapes of the grains. The science of materials may be defined as the science of everything which can be considered a bulk or a surface property, on any scale.

This definition is merely a logical one and has no relation to the prevalent subdivision of the science into physics, chemistry, metal-

lurgy, ceramics, crystallography, acoustics, etc. The boundaries between these different fields, though not unrelated to the logic of their subject matter, have been determined largely by the caprice of tradition and by the psychological bents of the types of minds which tend to predominate in each of the disciplines. The latter factor manifests itself, for example, in the tendency of the physicist to try to understand his material by means of a logically tight structure of quantitative relations, and to reject or ignore areas of knowledge which he cannot apprehend in this way. The chemist and the metallurgist, on the other hand, study many areas of the latter type, accumulate extensive data on them, and make useful generalizations about them, even though these generalizations have to be formulated in fuzzy or qualitative terms. Both approaches are true science, and both are necessary. The one is to be commended for its devotion to logic, the other for its willingness to apply scientific methods to the nasty problems as well as to the easy ones.

Within the limitations imposed by its coverage, the overall goal of this book has been to survey the science of materials from an impartial, interdisciplinary viewpoint. Some statistics on the panels who have authored the various chapters may be illuminating. According to the appellations by which the panel members choose to describe their occupations, a little over half of them are physicists, and the remainder are about evenly divided between chemists, or physical chemists, and metallurgists. About half of the chapters have been written by collaboration of men from all three disciplines; some have been monopolized by a single one. These variations have made for a diversity of technical language, but we hope that for the reader they will enhance the thrill of perspectives from different viewpoints.

The collaboration of different disciplines in the preparation of this book has brought home vividly to the participants—and we hope will do so to the readers—a fact which is widely known but still insufficiently acted upon: the science of materials needs, for its full flowering, both the techniques and the viewpoints of all the different traditional disciplines. In regard to techniques, the examples are legion. The physics of semiconductors made slow and uncertain progress until it was rescued by a tremendous effort of chemists and metallurgists in the direction of purification of materials. Much of surface chemistry remained confused and inexact until physicists' high-vacuum and electron-microscopic techniques were applied. Less obvious, but also important, is the stimulation of specialists in one discipline to pay attention to

problems which their natural prejudices incline them to ignore, or to make use of the viewpoint of a sister science.

3. GROWTH OF THE SCIENCE IN THE LAST HALF-CENTURY

The science of materials covers such a tremendous range of subjects that its diversity can only be appreciated after some study. Even the present volume does not touch by any means all of the subjects pertaining to crystalline solids. As a prelude to surveying the contents of the book it will be worthwhile first to survey the recent growth of the science of materials, and then to look in a little more detail at its various subdivisions and the distribution of research activity among them.

The growth of materials research in sheer quantity of publication can be roughly gaged from the statistics in Figure 1. Data from several abstract journals are shown to illustrate the difficulty of collecting meaningful statistics of this sort. There can be no doubt, however, that the literature of the field is enormous and rapidly expanding. Even on a logarithmic plot the overall curve seems to be concave upward.

Along with this snowballing growth in the volume of materials research, there has taken place a gradual but drastic change in its goals, in the nature of the questions which the scientist tries to answer. Namely, there has been an increasing trend toward the quantitative and definite aspects of understanding, as opposed to the qualitative and the approximate. In the first decade of this century, subatomic and quantum physics were in their infancy, and in their applications to solids the qualitative questions were the all-important ones: Do metals conduct because of free electrons? Are these the same entities which are found in vacuum tubes? Are atoms magnets? Today we worry over numerical values of effective masses of electrons in particular crystallographic directions in particular substances, and over the radial and angular distribution of the time-averaged magnetic moment of a particular atom in a particular crystal. In many areas of materials science an experimental study is considered deficient if the results cannot be quantitatively interpreted in terms of specific atomic mechanisms.

This stage of the science has been reached through the continued operation of two influences. One, of course, has been the tremendous progress in our knowledge of atoms and the laws of quantum physics, a knowledge which took an almost discontinuous spurt in the years close to 1925, but which has been growing

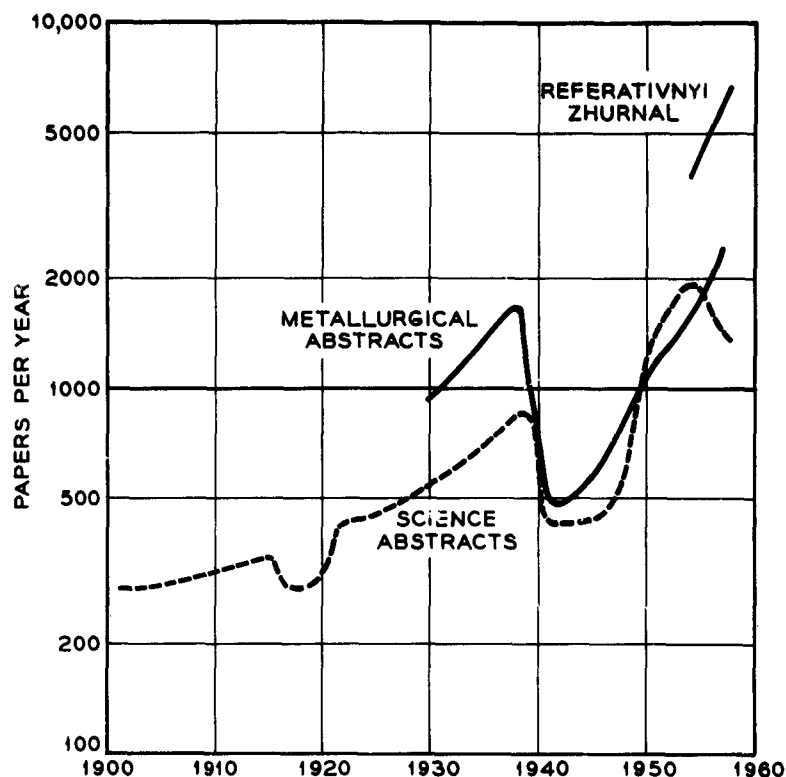


FIGURE 1.—Growth of the science of materials, as revealed by the numbers of papers listed in typical abstract journals. The curves for Science Abstracts (Physics) and Referativnyi Zhurnal (Fizika) were obtained by sampling certain issues of these journals and counting the numbers of papers pertaining to the pure science of solid state physics in the issues sampled; in adjacent years for which no samples were taken it was assumed that solid state and total papers were roughly proportional. The curve for Metallurgical Abstracts is simply the number of papers listed under the first three subject headings, namely, "Properties of Metals," "Properties of Alloys," and "Structure." As most of these entries represent pure science, and over half of them could be called solid state physics, the curve for Metallurgical Abstracts is more consistent with that for the Referativnyi Zhurnal than with that for Physics Abstracts. All curves have been smoothed.

steadily both before and since. The other factor has been the perfection of ever more powerful experimental tools: X-ray diffraction, cryogenics, microwaves, pile neutrons, techniques of purification and crystal growth, etc. Since World War II a not inconsiderable factor in the availability of these tools has been the economic importance of semiconductors and other highly specialized materials, which has enforced a much more intimate

cooperation of workers in different disciplines than would ever have been achieved otherwise, to the great benefit of all concerned.

A further aspect of modern experimental techniques deserves comment. Many of the properties of materials are structure sensitive—violently influenced by minute traces of impurities, or by occasional defects in an otherwise perfect crystalline arrangement of atoms. For decades much highly intelligent effort has been spent on the study of such properties, with meager results, because it has not been possible to isolate and identify the many different factors influencing a given property. In the last decade, however, enormous progress has been made, because control of the purity and perfection of materials has at last caught up with the demands of students of these properties, and because tools like electron microscopy and magnetic resonance have made possible an almost direct observation of atomic configurations. Herein lies the key to the increasing adequacy of scientific knowledge for the needs of technological development.

4. FIELDS AND DISTRIBUTION OF RESEARCH ACTIVITY

So far we have been concerned with the science of materials in toto. As a prelude to discussing the motivation of this book, and to encourage a charitable view of its shortcomings, we should now take a more detailed look at this science, with its fantastic diversity and uneven development. In this survey we shall map the pattern of present-day research activity from several different points of view: basic theoretical concepts, types of properties which are studied, kinds of materials on which the studies are made. All these factors are useful adjuncts for classifying the most vital things in materials research, namely, people and their relations to each other. Research is done by people, and knowledge accumulated by past research is useful only to the extent that people exist who will undertake to master it. The availability of scientific knowledge as a basis for technology depends on the existence of a body of experts to teach it. But more in need of stressing is the fact that, to an ever-increasing extent, modern science is jeopardized by fragmentation into narrower and narrower specialties, whose representatives are increasingly unable to communicate ideas which would be mutually helpful. We shall say more about this later on, but we must start, in this section, with an attempt to classify fields of materials research, and the specialists who pursue them.

Of the many different ways of subdividing current research on

materials, and the scientists who do it, the most useful is probably one based on the types of phenomena which can be studied. So let us consider first a classification on this basis. For our present purpose it will be best to use a classification which is a compromise between a purely logical system and one based on the recognizable categories of specialists, as these have grown up over the years. Even so, a list of fields would have to contain something like 40 or 50 entries to be sufficiently fine-grained to ensure a reasonable probability that an expert in a given one of the fields would have a reasonable familiarity with most parts of this field. Since the present section will have to consist mostly of tables, we shall try to make these as concise as possible by lumping together various fields which, though related, are sufficiently different to define quite distinct groups of research workers. With this sometimes brutal telescoping of categories, Table 1 lists the principal areas of present-day research in the science of crystalline materials. Where necessary, typical subfields belonging to each main heading have been indicated in parentheses. The tie lines give a rough picture of the interrelationships of the fields, a topic we shall take up in section 5. The circles opposite each major area depict the amount of current research activity in this area, and, for comparison, the amount of activity 20 years ago. From this comparison one can see the vast growth of some fields, the birth of completely new fields, and the steady, modest continuance of others. The proper scale of this growth is a little uncertain, but is probably to be obtained by comparing the 1938 circles with something in between the larger (*Referativnyi Zhurnal*) and smaller (*Physics Abstracts*) circles for 1958; this is because there is evidence that the criteria for inclusion in *Physics Abstracts* have narrowed somewhat since 1938.

One can also classify research work according to the types of substances which are studied. Throughout much of the science of materials it makes sense to separate metals, inorganic semiconductors, inorganic insulators, and organic crystals. The boundary between semiconductors and insulators is of course very nebulous, since many materials which are fairly good insulators at room temperature have interesting and even commercially important semiconducting properties at higher temperatures. In Table 2, which gives the distribution of research effort over these types of substances, we have classified the substance used in any investigation as a semiconductor if this substance as usually encountered has an appreciable (say $> 10^{-6}$ ohm-cm⁻¹) electronic conductivity at room temperature, or if the investigation was concerned specifi-

cally with semiconducting properties at a higher temperature. The comparison of 1958 with 1938 shows strikingly the effect of the post-transistor surge of interest in semiconductors. The other striking feature of Table 2, for both dates, is the very small amount of attention which has been given to the properties of organic crystalline materials. This neglect seems to have been due in part to the complex molecular structures of most such materials—modern solid-state scientists usually like to study the simplest things first—and in part to the artificial interdisciplinary barrier which causes physicists to consider organic compounds as lying in the chemists' domain, and chemists to consider plasticity, electrical conduction, and so forth, as belonging to the physicists' domain.

The people who do research in the science of materials are, usually, specialists in one or a very few of the fields or subfields listed in Table 1. Specialists who cover many properties of a particular material are rarer, although it often proves advantageous for students of a property most conspicuously manifested by a particular material to study some of its other properties as well. There is another type of specialization, however, of which we must take cognizance. This has to do with the division of research workers into theorists, who are usually as helpless as children in a laboratory, and experimentalists, who, though capable of mathematical and logical manipulations, spend too little of their time at these diversions to feel themselves in full command of current ideas. This dichotomy, though it has always been present to some extent, is much more pronounced today than it was a generation ago, because of the more elaborate development of experimental and theoretical techniques. In physics journals, where the proportion of theoretical papers runs higher than in the other sciences, about two-thirds of current publications in the solid-state field contain experimental results, while the remaining one-third, written usually by theorists but sometimes by experimentalists, are purely theoretical. The proportions are significantly different in different countries, however.

Most theoretical papers are addressed to a particular type of phenomenon, and so can be classified under one of the areas of phenomena which we enumerated in Table 2. A significant class of papers, however, deals with the basic building blocks of our picture of solids, and so is applicable to several or many types of phenomena. These papers, amounting to perhaps a few hundred a year, can be classified into several fairly well-defined groups, according to the type of concepts which they employ. Some—now-

adapts a small minority—treat a solid as a classical continuum, ignoring its atomic structure. A little less than half of the total treat solids as arrays of atoms, interacting by known or assumed forces and capable of all sorts of vibrational motions and departures from perfect crystalline order. The remainder, about half of the total, delve into the quantum-mechanical aspects of the states and motions of electrons in solids, or of the levels and interactions of electronic and nuclear spins.

For some purposes, it is useful to classify experimental scientists according to the specialized techniques with which they are familiar, regardless of the use to which these are put. However, only a fraction of current materials research uses techniques which are so specialized as to outweigh phenomena in their determining effect on the research pattern of an individual. As many of these will be discussed in Part XII of this book, we shall say no more on this subject here.

The final task of this section, to sketch the coverage of the present book in relation to the literature of the science, is best accomplished by returning to the property classification of Table 1. The chapter numbers after the entries show the way in which the various areas of solid-state phenomena have been divided among the parts of this book. It will be seen that the ratio of book space to extant literature varies over more than a factor of ten from one field to another. This variation in coverage does not necessarily imply a corresponding variation in the density of important ideas from one field to another; it is due rather to a superposition of accidental factors on top of a bias arising from an initial decision to give a slightly greater emphasis to the preparation and alteration of materials than to properties per se.

TABLE 1 (Pages 12 and 13)

Principal fields of the science of solid materials, classified according to the properties which can be studied. The amount of contemporary basic research activity in each field is indicated by the size of the circles opposite it: the area of the outer circle is proportional to the number of papers in the given field listed in the *Referativnyi Zhurnal (Fizika)* for 1958. Those of the middle and inner circles correspond to the number listed in *Physics Abstracts* for 1958 and 1938, respectively. In those fields in which there were more entries in *Physics Abstracts* for 1938 than in 1958, the middle circle is dashed but corresponds to the inner circle in other fields. The scale is such that the outer circle opposite "General theories," in the first row, corresponds to 370 papers.

The solid vertical lines are an attempt to depict the most closely related fields, an arrow from one field to another implying that the one contributes to the understanding of the other. The dashed lines show less vital relationships. After each entry, finally, are listed the parts of this book in which some of the subject matter of this area is surveyed. Listings with and without parentheses refer to chapters surveying, respectively, a minor or a major part of the area in question.

TABLE 1









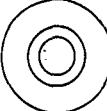

	General theories applicable to a variety of phenomena.	↑ ALL
	Crystal structure. Part VI.	↓ ALL
	Equilibrium energetics. (Binding energy, specific heat, phase equilibria, elasticity, expansion, etc.) Parts II, VI, VIII.	↓ MANY
	Dielectric polarization. (Dielectrics, piezoelectrics ferroelectrics, dielectric loss, etc.) Part IV.	
	Static and low-frequency magnetism. (Paramagnetism, antiferromagnetism, diamagnetism, ferromagnetism.) Part III.	↑
	Magnetic resonance. (Paramagnetic, antiferromagnetic, ferromagnetic, diamagnetic or cyclotron, and nuclear resonance.) Part III.	↓ MANY
	Thermal conduction in nonmetals. Part IV.	
	Superconductivity. Part III.	
	Electronic conduction. (Electrical and thermal conduction in metals, bulk conduction in semiconductors, electronic internal friction, certain aspects of photoconductivity, dielectric breakdown.) Part IV.	↑
	Optical phenomena. (Optical constants in the ultraviolet, visible, and infrared, luminescence, some aspects of photoconductivity and photochemical processes.) Part IV.	↓

TABLE 1—Continued

○	Nonelectrical properties of grain and interphase boundaries. Part IX.	
⊙	Miscellaneous electrical properties of surfaces and interfaces. (Contact resistance, rectification, contact potential, surface conduction, etc.) Part IX.	⋮
⊙	Electron emission, etc. (Thermionic, photoelectric, secondary, and field emission, reflection of electrons from surfaces, exoemission.) Part IX.	
⊙	Fine-particle and thin-film systems.	↕ MANY
●	Friction and adhesion. Part IX.	
⊙	Atomic arrangements and motions on surfaces. (Structure and energy of surfaces, adsorption, migration, catalysis, crystal growth, etc.) Part IX.	↑ ALL SURFACE
⊙	Plasticity, rupture, and dislocation phenomena. Part VIII.	↑
○	Kinetics of phase changes. Parts VI, VII.	↑
⊙	Atom movements. (Diffusion, electrolytic conduction, alloy ordering, annealing of point defects, some types of internal friction, etc.) Part V.	↑
⊙	Passage of particles through solids. (Diffraction and elastic or inelastic scattering of electrons and neutrons, radiation damage, etc.) Part XI.	↑
⊙	X-Rays (Effects not covered under "crystal structure" above.)	⋮

TABLE 2

Distribution of research papers on properties of crystalline solids among different types of materials. Sources: Referativnyi Zhurnal (Fizika) 1958, and Science Abstracts (Physics) 1958.

Type of Material	Fraction of Papers	
	1958 Percent	1958 Percent
Metals	52	46
Inorganic Semiconductors	9	19
Inorganic Insulators	34	28
Organic Crystals	5	7

5. INTERRELATIONS OF DIFFERENT FIELDS

The term "science of materials," which we have found it convenient to adopt, carries the connotation that a central goal of this science is the understanding and control of the way in which differences between one material and another affect their various properties. While studies of any property on different materials automatically contribute to this goal, its full realization depends on understanding the interrelationships of different properties. Let us therefore take a brief look at the contemporary pattern of interactions between studies of one type of property and another.

In Table 1 an attempt has been made to show, by means of tie lines, which fields are notably interrelated in present-day research. A full line signifies a very close relationship, a dashed line a less close one. Some fields, such as crystal structure or magnetic resonance, are related to so many of the others that the lines have been omitted and replaced by the word "all" or "many." Arrows have been placed on the full lines in an attempt to distinguish two-way from predominantly one-way relationships: often one field provides a tool for studying a second field, but the second has little reaction on the first; in such cases only one arrow has been drawn, pointing toward the second field. The drawing of the lines is of course very much a matter of opinion, and they should not be taken too seriously. Moreover, they leave unspecified many important details: often one subfield of a field listed in the table is closely related to another field, while other subfields are not. Despite its crudity, however, the pattern of relation-lines will suffice to illustrate several important points.

Consider first the role of theoretical concepts. If the study of one phenomenon is related to the study of another, it is usually

because the theoretical concepts underlying the two fields are in part the same. For example, concepts having to do with the electronic states of crystals are important for most of the fields whose location in the table is between dielectric polarization and X-rays; correspondingly, there are a number of lines connecting these fields. Radiation damage and the fields from atom movements through fine-particle and thin-film systems, on the other hand, involve mainly the conception of atoms as displaceable units which exert forces on each other; these fields again form an interconnected complex. One would expect that the number and closeness of interrelationships of this sort would be proportional to the degree of refinement of the theories which provide the connections; in the history of the various fields this has indeed proved to be so. Solid-state physics papers of the first decade of this century show much less interrelation of different fields than occurs today. Concomitantly, in this period there were almost no papers on what we have called "general theory"; practically all theoretical papers pertained to a particular type of phenomenon. Even as late as 1925, there was very little work on broad general topics in solid-state theory. The development of the quantum theory of solids, and of concepts like lattice vacancies and dislocations, has made the era since 1930 one of rapidly increasing interactions between fields.

However, the existence of common theoretical concepts is not the only source of relation-lines in Table 1. For example, the various types of surface phenomena form an interrelated complex, and while many of the relations have to do with atomic concepts, many of them have to do merely with the fact that all studies of free surfaces face one major problem together, the experimental problem of obtaining a reproducible surface of determinable atomic constitution. Only with modern high-vacuum techniques is this problem beginning to be solved, and phenomena from many of the different fields that relate to surfaces have contributed to the solution.

Some fields underlie most of the others. The field of general theories falls in this category by definition. Crystal structure is also fundamental, as it plays a major role both in determining electronic states and in phenomena involving dislocations, atom movements, and surfaces. Similarly, the equilibrium energetics of any material are basic to many of its other properties.

Finally, certain fields have provided powerful tools with which the problems of other fields can be investigated. This type of interrelationship usually depends very much on the existence of a

common pool of theory; however, the field receiving the tool often benefits much more from the interaction than the field which supplies it. The outstanding example of the present day is magnetic resonance, which has provided a wealth of information about free electrons, impurity centers, etc., and which thereby has solved many problems in fields such as electronic conduction, optical phenomena, and radiation damage. Other fields, such as optical phenomena, electron diffraction, and electron emission, have also provided important tools.

6. THE PROBLEM OF PERSPECTIVE

The mere enumeration of such a wide variety of fields as are listed in Table 1 should suffice to illustrate how difficult it is for one man to have any adequate perspective of more than a few of them. Interrogations of randomly selected scientists seem to give a picture of the typical active experimental research worker as having a fairly good knowledge of his own subfield (one of the items in parentheses in Table 1), a fair awareness of important trends in several related subfields, and only a sketchy idea of what is going on in all the rest. In other words, the scale of subdivision which we have used in Table 1 is of the same order of magnitude as the range of perspective of the more active experimental scientists. Theorists, as one might expect, tend to have a rather broader perspective, but one which rarely extends over more than a fraction of the fields that comprise the science of materials.

Of course, no one tries to read the full literature of any but the few narrow fields in which he is working. But a few decades ago one could almost always get a fairly complete, though not quite up-to-date, picture of any other field by referring to textbooks, handbooks, review articles, etc. Today, despite an ever-increasing number of books and reviews, this is getting to be less and less possible. Although the primary cause for this state of affairs is the sheer bulk of the literature, this bulk needs to be considered in relation to certain other factors, which we shall now discuss.

Let us look for a moment at two clearly definable fields, picked arbitrarily from the entries of Table 1: electron theory of solids (including its applications to specific phenomena) and electronic conduction. The former, a theoretical field, was thoroughly reviewed in volume XXIV-2 of the Geiger-Scheel *Handbuch der Physik* (Julius Springer, Berlin, 1933), mainly in the article of Sommerfeld and Bethe, partly in one or two of the other articles. The latter, primarily an experimental field, was thoroughly reviewed by Meissner, Kohler, and Reddemann in volume XI-2 of

the Wien-Harms *Handbuch der Experimentalphysik* (Akademische Verlagsgesellschaft, Leipzig, 1935). In each of these reviews, mention was made of something like half or two-thirds of the existing research papers in the field, exclusive of outdated ones; there is good reason to believe that very few ideas of any significance were missed. In these reviews the ratio of pages of review to the number of pages in the original research papers cited (in the field under review, not in peripheral fields) was about 1:8 for the theoretical review, and 1:20 for the experimental. Application of a similar ratio to the literature of value today would suggest several thousand pages of review for each field.

Since this is clearly far too much for a single author or small team of authors to undertake, the only hope of getting adequate review coverage of today's literature would be either to have more articles with each covering a much smaller amount of ground, or to have very comprehensive treatises prepared by a vast coordinated effort of many collaborators. But both alternatives bog down. For example, if one tries to cover electron theory by independent articles on subjects like "electron interaction in metals," "determination of the Fermi surface in metals," etc., one begins to find that the width of the area covered by each review is of the same order as the range of uncertainty in defining the ground which the review covers. Thus the totality of available review articles constitutes a random spatter over the ground, and few articles have comprehensive coverage over the field implied by their titles. The coordinated-effort idea, on the other hand, is so unwieldy that it has scarcely been tried; it would probably lead to such delays as to make the resulting comprehensive article out of date before it appeared.

It is clear that another level must be inserted into the review literature, one which deals with broad trends and major ideas over a large area, and does so as objectively and all-inclusively as possible. At this level one must give up attempts to develop a tight logical structure in which each conclusion is thoroughly buttressed by logic or data; the goal must be breadth of perspective, with facts and ideas merely stated, though in such a way as to show their origins, their interplay, and their uncertainties. This book is an attempt in this direction: it is a "perspective," designed to orient the research planner and the pedagogue, and to aid all who need to map their way through the review literature. Its method of production has been intermediate between that of a tightly coordinated project and that of a compilation of independent articles. A moderate amount of overall direction has been

exercised by the committee in charge of the project, and within each chapter the panel members assigned to it have collaborated, sometimes loosely, sometimes very closely. Thus, the resulting perspective usually reflects the collective view of a number of experts. It undoubtedly falls far short of the ideal both in balance and in depth; however, the very shortcomings of this book, arising as they do from the imperfect perspective of the committee and panel members, provide the strongest possible argument that such a book is needed.

Part II

COHESIVE PROPERTIES OF SOLIDS

Chairman

Prof. HARVEY BROOKS
Harvard University
Cambridge, Mass.

Members

Dr. J. O. BETTERTON
Oak Ridge National Laboratory
Oak Ridge, Tenn.

Dr. L. GUTTMAN
General Electric Co.
Schenectady, N. Y.

Dr. J. FRIEDEL
Université de Paris
Paris, France

Dr. E. W. HART
General Electric Co.
Schenectady, N. Y.

Dr. JOHN R. REITZ
Case Institute of Technology
Cleveland, Ohio

CONTENTS

	Page
1. INTRODUCTION	21
2. RIGOROUS THEORY OF COHESION FOR PURE ELEMENTS AND COMPOUNDS	22
3. ELASTIC CONSTANTS AND INTERATOMIC FORCES	28
3.1. Individual Contributions to the Elastic Constants	29
3.2. Ion-Ion Repulsion	31
3.3. Force Constants and Vibration Spectra	33
4. QUANTUM BASIS OF CRYSTAL CHEMISTRY	33
4.1. Hume-Rothery Rules and Related Ideas	34
4.2. Coordination Number and Microscopic Symmetry in Crystals	36
4.3. Relative Stability of FCC, BCC and HCP Structures in Metals	37
4.4. Special Examples	38
4.4.1. Mooser-Pearson Semiconductor Rules	38
4.4.2. The Refractory Hard Metals	39
5. THEORY OF THE COHESIVE PROPERTIES OF ALLOYS	40
5.1. Fundamental Methods	40
5.1.1. Screening	40

5.1.2. Exclusion Principle and Correlation.....	41
5.1.3. Electronic Contribution to Binding.....	41
5.1.4. Ion Core Repulsion and Lattice Relaxation.....	42
5.1.5. Critique	42
5.2. Models and Semi-Empirical Methods.....	43
5.2.1. The Rigid-Band Model.....	43
5.2.2. The Elastic Strain Model.....	44
5.2.3. Statistical Models	44
5.2.4. Critique	44
6. METHODS OF MEASUREMENT OF COHESIVE PROPERTIES	45
6.1. Cohesive Properties of Pure Crystals	45
6.1.1. The Cohesive Energy.....	45
6.1.2. Microscopic Elastic Constants	47
6.1.3. Present Status of Knowledge of Cohesive Properties.....	48
6.2. Cohesive Properties of Alloys and Other Multicomponent Systems	49
6.2.1. The Cohesive Energy	49
6.2.2. Microscopic Elastic Constants.....	50
6.2.3. Present Status of Knowledge of Cohesive Properties of Alloys	50
7. REFERENCES	51
8. APPENDIX: COHESION OF METALLIC ALLOYS—A SHORT THEORETICAL SURVEY by J. Friedel.....	52
8.1. Main Factors Affecting the Cohesion of Alloys	52
8.2. Usual Approximations	54
8.2.1. Independent-Electron Model	54
8.2.2. Perturbation Methods	55
8.2.3. Free-Electron Approximations	59
8.2.4. Tight-Binding Approximations	62
8.2.5. Better One-Electron Approximations	62
8.2.6. Size Effects	63
8.3. Energies of Primary Solid Solutions.....	65
8.3.1. General Considerations	65
8.3.2. Solution Entropy	66
8.3.3. Interactions Between Solute Atoms	67
8.3.4. Stability Range of Primary Solid Solutions	68
8.3.5. Allotropic Transformations	69
8.4. Energies of Intermetallic Phases	69
8.4.1. Ionic and Covalent Factors.....	69
8.4.2. Size Factor	70
8.4.3. Electron Compounds	70
8.5. Lattice Constants	71
8.5.1. Infinite Dilution	71
8.5.2. Anomalies for Zone Overlaps.....	71
8.6. Elastic Constants	71
8.6.1. Infinite Dilutions	71
8.6.2. Anomalies for Zone Overlaps	72
8.7. Instructive Experiments	72
8.8. References	73

COHESIVE PROPERTIES OF SOLIDS

1. INTRODUCTION

The development of the quantum theory of the hydrogen molecule pointed the way to a general understanding of the origin of binding in solids. However, except for the purely classical theory of ionic crystals, the first quantitative attempt to compute the cohesive energy of a solid from first principles was the treatment of metallic sodium by Wigner and Seitz (1934). Almost all subsequent treatments of cohesion in metals have been based on this approach, and it still appears that the calculation of Wigner and Seitz contained the most useful breakdown of the various contributions to cohesion in metals having rare-gas type ion cores. Despite the early success of this theory it has not been very useful in dealing with properties that depend explicitly on crystal structure. As a consequence, most of the progress in the crystal chemistry of metals as well as non-metals has resulted from the application of qualitative rules derived from valence theory—rules which have never been fully justified from first principles. Indeed the two points of view, band theory and valence bond theory, have tended to develop into divergent schools of thought, with a different language and a somewhat heated partisanship. The rather sterile nature of this controversy has been appreciated in recent years.

Energy differences between allotropic modifications of most solids are less than 3 kcal. per mole, which is far less than the uncertainties in the fundamental theoretical calculations. Thus, theories of cohesion based on band theory, i.e., those based essentially on the Wigner-Seitz method, have so far been able to handle only contributions to the cohesive energy which depend mainly on the atomic volume. On the other hand, the valence bond picture cannot be precisely related to the crystal wave functions, or any actual approximation thereto, and is therefore little more than a systematic translation of empirical observations into seemingly more theoretical language. Nevertheless, the valence bond approach has proved useful in guiding the search for new materials and even in making qualitative predictions of some of their properties. Attempts to refine valence bond theories for solids are not likely to be very rewarding, however. Although not wholly satisfying as scientific theories, valence bond rules possess a surprising simplicity which contrasts with the apparent complexity and arbitrariness in the calculations from first principles, sug-

gesting that these rules may have their origin in more fundamental considerations which are independent of specific models.

In this report we have not attempted to reconcile the various viewpoints but have tried to ensure that both are adequately represented. We have also emphasized the importance of trying to justify valence bond theory, as applied to crystals, in quantum mechanical terms. As indicated above, it may be that one of the most fruitful lines of development for the future will lie in attempts to create theories and principles which are independent of the traditional models and approximations.

In section 2 we review briefly the status of the rigorous theory of cohesion as applied to various solid types. In section 3 we take up the theory of elastic constants and interatomic forces, pointing out that an understanding of the origin of elastic constants may contribute towards a theory of the stability of phases, which is considered next in section 4. Here we discuss what little can be said about crystal chemistry from a fundamental standpoint, especially as regards the prediction of the relative stability of phases of pure materials. In section 5, a review of the theory of alloys is given with emphasis on the electronic origin of the thermodynamic parameters which govern the phase diagram. Finally, in section 6, we consider methods of measurement of the cohesive properties of pure materials and alloys and review the basic thermodynamic principles involved in phase diagrams and cohesion.

2. RIGOROUS THEORY OF COHESION FOR PURE ELEMENTS AND COMPOUNDS

There are several recent reviews of the theory of cohesion in simple solids. Mention may be made especially of the papers by Wigner and Seitz (1955R), Löwdin (1956R), Callaway (1958R), Brooks (1958R), Cohen and Heine (1958R), and Pines (1955R). These reviews are concerned more or less with the problem of cohesion from a fundamental quantum standpoint. Several other reviews have been published which deal in more detail with the technique of energy band calculations; this, however, represents only a part of the cohesion problem.

A survey of the existing experimental data on the cohesive energies of pure elements reveals many striking regularities (Wigner and Seitz, 1955R). For example, with some systematic exceptions, the binding energy per nominal valence electron varies by less than a factor of two throughout the periodic table. The outstanding exception is the monovalent noble metals where some

binding is believed to arise from the electrons in the closed d shells which overlap each other in these metals (Brooks, 1958R). The binding energy per electron decreases systematically on moving down a given column of the periodic table, except towards the middle of the transition group series where an increase is observed. There is not much difference between the binding energy per electron in metals and that per nominal bond in covalent crystals. These regularities have so far received only qualitative explanations.

Attempts at a rigorous calculation of cohesive energy from first principles have been restricted to the simplest metals and ionic crystals. For metals, the computation begins with an attempted solution of the Hartree or Hartree-Fock equations. The basis wave functions are representations of the fundamental translation group of the crystal, that is to say, the square of the wave function has the translational periodicity of the crystal. This portion of the problem consists of two parts: (1) the determination of a suitable "one-electron potential, and (2) the solution of Schrödinger's equation in this periodic potential. Much more progress has been made with the second problem than with the first. A number of new mathematical techniques, such as the orthogonalized plane-wave (OPW) method (Woodruff, 1957) the augmented plane-wave method (Reitz, 1955R), and the Kohn-Rostoker (1954) variational method, have been applied successfully to the solution of the periodic potential problem with the aid of high-speed electronic computers.

Although some difficulties have arisen with convergence, associated with the finite capacity of the machines, the problem may nevertheless be said to have been solved in principle. Except possibly by using the Kohn-Rostoker method, it is not yet feasible to solve the Schrödinger equation at points other than symmetry points within the Brillouin zone, although the variation of energy with crystal momentum may be determined by applying a perturbation theory at nearby points. With the OPW method, the least difficulty is encountered in using a realistic crystal potential, but the convergence is least reliable and requires the greatest number of terms in the secular determinant. Nevertheless, it is safe to say that at present the calculational accuracy of solving the Schrödinger equation exceeds the accuracy with which the crystal potential can be determined in practice.

The determination of the crystal potential can be resolved quite naturally into two major problems: (1) estimation of the potential due to the ion cores, including the effects of exchange and correlation between the valence electrons and the ion cores, and

(2) estimation of the Coulomb and exchange potentials arising from the other valence electrons. There are also secondary corrections due to the modification of the ion-core potential by the valence electrons in the solid; these have scarcely ever been considered, although they are mentioned by Seitz (1935).

A satisfactory solution does not even exist to the first problem. Free atom Hartree-Fock potentials, including the exchange operator, provide a reasonable approximation for elements up to about aluminum, but, for heavier elements, polarization of the core by the valence electron becomes increasingly important and relativistic effects such as spin-orbit interaction must also be taken into account. This difficulty can be partly circumvented by the use of the quantum defect method (Ham, 1955), in which the properties of the solution to the Schrödinger equation outside the ion core are deduced from experimental values of the spectroscopic terms for the free atom, which automatically incorporate the effects mentioned. Even in this method, however, the polarization of the core by a given valence electron is screened by the presence of the other valence electrons, and this effect cannot be taken into account accurately in the calculation, partly because of the inadequacies of our theory of the correlation energy between valence electrons. The quantum defect approach is thus difficult to justify rigorously from first principles when core polarization is significant (Brooks and Ham, 1958). It is believed, however, that further progress is possible along these general lines.

The solution to the second part of the crystal potential problem is even less satisfactory. The Coulomb potential due to the other valence electrons can be computed by standard electrostatic methods, although it is often omitted from the calculations in practice. The computation of the exchange potential, while straightforward in principle, is extremely complicated in practice; it is often treated by using the so-called Slater approximation in which the exchange effect is replaced by a potential energy which varies as the cube root of the local charge density. This simple approximation is probably incorrect, but it is difficult to estimate the error introduced in any actual calculation. A more rigorous treatment of the exchange energy than has yet been carried out is doubtlessly possible, but the incentives for the necessary labor are not great, because of the impossibility of treating the equally important correlation energy on a comparably rigorous basis. Exchange can always be treated in principle by introducing a one-electron potential which is nonlocal in character, but it is not clear

that correlation between valence electrons can be handled by this simple generalization of the one-electron theory.

Having obtained the best possible one-electron energies from the solution of a Hartree-Fock equation, it is still necessary to compute the effects of correlation, mainly between electrons of antiparallel spin, in order to determine the total energy of the system. This problem has been handled in detail only for a free-electron gas (Pines, 1955R), and even here the solution is far from satisfactory. Good approximations can be made for the limits of high electron density (the so-called "random phase" approximation) and low electron density (the "lattice" approximation), but a satisfactory theory does not exist for the intermediate density range, which corresponds to the actual electron density in metals.

In the intermediate density range, the total energy can be estimated by interpolation between the two limiting or extreme cases, but with the alkali metals, for example, the estimated accuracy amounts to about 20 percent of the entire cohesion. Nozieres and Pines (1958) have discussed this problem in relation to real metals, and although a general formalism is available, it has so far not proved possible to handle such systems in practice. In most calculations it is assumed that correlation and exchange energies are sensitive only to atomic volume and are insensitive to structure, that is, to the geometry of the unit cell. This may not be sufficiently accurate when considering the small energy differences which govern phase stability, or when discussing such properties as the elastic constants. Recent advances in the mathematical techniques for handling many-body problems (Martin and Schwinger, 1959), give promise that an improved theory of correlation in the intermediate density range may be forthcoming, as well as an understanding of other electronic properties which are less sensitive to the specific approximations of the band picture than is usually believed.

Another approach to the correlation energy problem is to seek alternative solutions to the Hartree-Fock equations for which the wave functions are not representations of the translation group but rather have periodicities of a "super lattice" type. Some preliminary work along these lines has been carried out by Löwdin (1956R), and more recently by McIrvine and Overhauser (1959), and this work shows some promise. The solutions obtained are such that the wave functions for different spin directions have different spatial distributions, in response to which electrons of antiparallel spin tend to stay apart automatically; in this way, some correlation is incorporated into the one-electron wave func-

tions. This approach may offer some hope of taking into account the actual crystal structure, but experience in atomic and molecular calculations using this method suggests that only a part of the correlation energy can be accounted for with wave functions which do not contain the interelectronic distance explicitly. For example, the most accurate molecular calculations for diatomic molecules, using Hartree-Fock wave functions and configuration interaction, but without explicit introduction of interelectronic distances, give errors in the total energy of the system of the order of 1 rydberg (Allen and Karo, 1960).

A number of attempts have been made to treat cohesion in crystals starting from the wave functions of the free atoms. For metals, semi-metals, and semiconductors, the so-called "tight binding" or LCAO method has been used (Reitz, 1955R), corresponding to the Hund-Mulliken method for molecules. This method has the potential advantage that some of the correlation energy effects which arise in polyvalent metals can be made to cancel insofar as they are also present in the free atom. The disadvantage of the method is that nonorthogonality of wave functions on different atoms (including the core wave functions) is difficult to handle, and when it is handled properly the resulting secular equations become exceedingly complex. Recently, a promising approach to handling the problem of nonorthogonality has been devised by Stern (1959), who has applied his modified LCAO method to a discussion of cohesion in iron. There may be some hope of using the same method to generate wave functions having different spatial distributions for different spins (as in Löwdin's method) so as to take into account correlation.

The most satisfactory cohesive energy calculation is probably that carried out for the alkali halides by Löwdin (1956R), using determinants of atomic wave functions and taking into account all of the nonorthogonality effects. Such a calculation is manageable for the alkali halides because the electrons are in closed shells. Even here, however, the agreement with experiment is only fair, and the calculation is not sufficiently accurate to predict the structure reliably.

The next most satisfactory calculation is that for the alkali metals (Brooks, 1958R; Ham, 1955), using the quantum defect method. These calculations, however, have so far not been carried out using the most realistic possible potential, and no effort has been made to take into account the influence of the lattice on the Coulomb, exchange, and correlation energies. Furthermore, the effect of core polarization has not been adequately handled be-

cause of the problem of screening by valence electrons mentioned previously. Thus, although present results are in apparent agreement with experiment, too many effects have been neglected to attach much significance to this agreement. In the absence of an improved theory for the correlation energy, further refinement of this type of calculation is probably not justified for the present.

Very few calculations have been attempted for other metals. It is possible to obtain good order of magnitude agreement for the cohesive energies and lattice constants of polyvalent metals using the quantum defect method combined with some drastic simplifying assumptions due to Raimes (1952). In this method only the energy of the lowest state of the valence electron is computed, and the electrons are otherwise treated as being free.

It is felt that insofar as the cohesive energy is concerned, comparisons between theory and experiment are of little significance except when the sources of error in the theory can be fairly accurately estimated.

A discussion of the rigorous theory of cohesion cannot be separated from the more general problem of the electronic structure of solids and the nature of wave functions and energy levels in solids. Recent general considerations of the many-body theory of solids indicate that one-electron concepts, such as the Fermi surface and the motion of carriers in external electric and magnetic fields or in the fields of impurities in semiconductors, may have a validity which extends far beyond the one-electron picture or the Hartree-Fock approximation. On the other hand, in seeking a quantitative comparison between theory and experiment, it is far from certain whether the Hartree-Fock energies give a satisfactory description of the band structure as deduced from measurements of electronic properties such as cyclotron resonance (Kip, 1960) and the de Haas-van Alphen effect (Shoenberg, 1960).

For example, there are several experimental indications that the Fermi surface in lithium touches the zone boundary (Cohen and Heine, 1958R), whereas the best theoretical work definitely predicts a nearly spherical Fermi surface, with an anisotropy much less than indicated by magneto-resistance measurements. There are also discrepancies between the predicted and experimentally determined band structure for aluminum (Heine, 1957), although here the interpretation of the experiments is not yet unambiguous (Harrison, 1959). It may well be that one of the most fruitful lines for future development will be to establish general methods for correlating empirical determinations of electronic properties with cohesive properties. For example, once the details of the

shape of the Fermi surface in aluminum are understood through cyclotron resonance experiments, it should be possible to develop a theory for the elastic constants of aluminum and even to predict phase boundaries of aluminum alloys much more reliably than could be achieved in a purely theoretical calculation from first principles.

In this connection, the development and exploration of new techniques for studying the band structure of metals assumes a special importance. The most powerful techniques are those such as cyclotron resonance, the de Haas-van Alphen effect, ultrasonic attenuation, anomalous skin effect, etc., (Harrison and Webb, 1960R) which permit a more or less point-by-point exploration of the properties of the Fermi surface. Studies of electronic transport properties, such as galvanomagnetic effects and thermoelectric power, and of magnetic and thermal properties, especially the electronic specific heat, form a valuable adjunct to the more detailed methods; however, measurements of this type are much more useful in verifying or fixing some of the parameters in a fairly well-established model than in deducing the band structure directly. As has already proved true in the case of semiconductors, it is felt that a combined theoretical and experimental attack on the problem of electronic structure will prove much more fruitful than a purely theoretical approach; nevertheless, we feel that theoretical work should continue, especially to help establish the limits of validity of the one-electron theory. Major efforts should be directed towards establishing those general theoretical conclusions which are independent of specific approximations and models.

3. ELASTIC CONSTANTS AND INTERATOMIC FORCES

The elastic constants of a crystal are defined in terms of the second derivatives of the free energy density of the crystal with respect to the strain. Because the various atomic interactions which contribute to cohesion depend differently upon atomic separations, the double differentiation has the effect of changing the relative importance of these interactions compared to the role they play in determining the cohesive energy itself. Thus, the approximations which suffice for a quantitative theory of cohesion may be inadequate for a quantitative theory of the elastic constants. On the other hand, a detailed comparison between theory and experiment as applied to the elastic constants may provide additional information about interatomic forces and electronic structure beyond that which can be obtained from a consideration of the cohesive energy alone.

A fundamental theory of elastic constants requires a knowledge of the change in crystal energy with strain. The direct approach, in which a rigorous calculation is made of the electronic energy levels in a series of strained configurations, is satisfactory for the bulk modulus provided the crystal has sufficiently high symmetry, and this approach has been used in calculating the shear constants of the alkali metals. In most cases, however, more drastic approximations are necessary. The approximations usually made are the following:

(1) The energy of the crystal is written as the sum of various individual contributions, and the change of each such contribution with strain is then treated independently.

(2) The variation of these energy terms with strain is examined on the basis of a simplified model which usually involves several undetermined parameters. These unknown parameters may then be evaluated from other experimental data or chosen to make the energy term agree with a rigorous calculation of the same term in the energy of the unstrained crystal.

(3) Exchange and correlation energies are treated as being dependent primarily upon the atomic volume and the effect of the crystal geometry on these terms is neglected. Thus, it is assumed that these terms do not contribute to the shear moduli corresponding to strains at which the crystal volume does not change. It must be emphasized that this assumption has never been properly justified by a fundamental calculation, although it seems to be a reasonable approximation for the alkali metals.

(4) Calculations of elastic constants are usually made for the absolute zero of temperature. Thus, the electrons are assumed to populate their lowest energy states, the appropriate lattice constant at 0°K is used, and entropy effects are neglected. This restriction is not essential, but in most cases the added complication introduced by employing a finite temperature is not justified because of the many other approximations involved in the calculation.

3.1. Individual Contributions to the Elastic Constants

The electrostatic contribution to the elastic shear constants of metallic crystals may be calculated on a purely classical basis using the electronic charge density distribution for the crystal (Huntington, 1958R). The electrostatic contribution to the shear constants may be computed on the assumption that the electron density around each ion remains spherically symmetric and centered on the ion, whereas the electron density, $\rho(r_i)$, near the

boundary of the Wigner-Seitz cell stays constant during the shear. This procedure is equivalent to the shearing of a lattice of point ions in a uniform "sea" of negative charge, $\rho(r_s)$, a procedure used by Fuchs to calculate the elastic constants of the alkali metals.

The contribution to the elastic constants arising from the overlap of ion cores is usually derived from repulsive energy interactions of the form

$$W_{12}(r) = Ae^{-r/\rho}. \quad (3.1)$$

This contribution is discussed in more detail in section 3.2.

It may be shown that the electrostatic contribution to the elastic shear constants is proportional to the square of the effective charge density at the boundary of the Wigner-Seitz cell. For real crystals this differs from the value which would be computed on the assumption of a uniform distribution of electronic charge. The difference may be computed theoretically and checked against the experimental results with fair agreement. The charge density of the cell surface also contributes to the dependence of shear constants on pressure which, it appears, is considerably greater than would be expected from simple theory (Smith, 1959). Measurement of the elastic constants as a function of pressure may, in fact, provide a unique means for relating the elastic constants to the electronic structure.

In simple metals having spherical electron energy surfaces, the Fermi energy should not contribute to the elastic shear constants and should contribute to the bulk modulus only through the direct compression of the "electron gas." This is the situation usually assumed to govern the behavior of the monovalent metals. However, recent experiments have shown that the Fermi surface is not spherical in copper, silver, and gold (Harrison and Webb, 1960R). Furthermore, the sign of the thermoelectric power in lithium is incorrectly given by the simple model, and it has been suggested that this is also due to a nonspherical Fermi surface. Thus, even for the monovalent metals, the Fermi energy probably makes some contribution to the shear constants, and experimental deviations from the simple electrostatic theory have been verified for lithium. For multivalent metals, the Fermi surface is usually complicated in shape, coming into contact with the Brillouin zone in several places.

Leigh (1951) has developed a phenomenological theory for computing the contribution of the Fermi energy to the shear constants in such cases. His model is based primarily on geometric considerations and contains a number of undetermined parameters,

but it has been found possible to fit his theory to the observed elastic constants when sufficient data are available. Substantial progress in this area may be expected in the future, particularly by making use of knowledge of the band structure derived from cyclotron resonance and other techniques to predict the elastic constants, and then comparing the resultant predictions with experiment.

3.2. Ion-Ion Repulsion

The ion-core repulsion represents an important contribution to the elastic constants of the noble metals and the transition metals. In metals of both groups, the dominant interactions are those between d shells. The simple Born-Mayer type of interaction mentioned in section 3.1 appears to be inadequate and, indeed, has only a weak theoretical basis. This form of interaction has been derived for two interacting closed p shells (or closed s shells) in the limit in which electron correlations are neglected, whereas at large ion separation it is known that these correlations give rise to a van der Waals interaction.

The general theory of closed-shell interactions is based on the theoretical work of Slater, of Born and Mayer (Seitz, 1940), and of Löwdin (1948). These authors investigated the interaction of closed s shells or closed p shells within the framework of first-order perturbation theory and found an exponential repulsion of the form given in Eq. (3.1), with a coefficient roughly proportional to the number of electron-pair interactions. However, the attractive term, the van der Waals interaction, is derived from second-order perturbation theory for the limiting case where core overlap is negligible. The counterpart of the van der Waals interaction for the case where the "shells" overlap has not been studied in detail but may have a profound effect on the overall ion-ion interaction. This is a field of investigation that has been comparatively neglected recently but could become quite fruitful with newly developed techniques for molecular calculations.

Mott has suggested that a substantial part of the cohesion in the noble metals may arise from an attractive interaction between the supposedly closed d shells. Some confirmation of this idea comes from the work of Daniels and Smith (1958), who have shown that it is impossible to account for the elastic constants of the noble metals, as well as their pressure coefficients, by assuming only an electrostatic contribution and a contribution deriving entirely from centrally acting repulsive interactions. In the transition metals the presence of unfilled d shells should produce even

stronger correlation effects. Thus, here again we find an area where the lack of an adequate method for treating many-body problems has held back theoretical progress.

In the alkali halides the situation is somewhat clearer, for here we are dealing with interactions between closed p shells which are much "harder," i.e., less polarizable, than d shells. The major contribution to the elastic constants comes from the repulsion of the closed shells, and the elastic constants have been calculated successfully for a number of different approximations. The most fundamental work is that of Löwdin (1948), who used the detailed quantum mechanical approach based on the Heitler-London method. Löwdin found that the principal contribution from the interacting ion cores is a centrally-acting repulsive interaction of the Born-Mayer type, but in addition certain three-center integrals appear. These integrals give rise to multibody forces which are the origin of the observed deviations from the Cauchy relations. Löwdin's work predicts the correct sign and approximately the correct order of magnitude for these deviations.

For the crystalline rare gases, predictions based on exponential repulsions and van der Waals interactions are in good agreement with observed elastic properties (Huntington, 1958R), although experimental results are presently available only on polycrystalline materials. It would be expected that the theoretical model would be rather good for these materials, and experimental studies, both on the elastic properties and on the lattice vibration spectra of solidified gases, though very difficult, should provide a valuable basis for comparison between theory and experiment.

The interaction potential between closed shells (atoms or ions) can also be approached directly by determining the virial coefficients of compressed gases or by direct atomic scattering experiments (Hirschfelder, Curtiss, and Byrd, 1954; Amdur, 1958). It should be mentioned that the Buckingham-Corner potential appears adequate for neon and argon; calculated and experimentally determined second virial coefficients agree very closely over a 600° temperature range. However, virial coefficient determinations do not give much information about the region where substantial overlap occurs, which is important for elastic properties; atomic beam experiments, on the other hand, do give such direct information and should be especially valuable.

The threshold energies for producing defects in radiation-damaged solids, and particularly the surface sputtering threshold and directional effects resulting from ion bombardment, are quite sensitive to the form of the repulsive interaction between ion cores.

Sputtering experiments may possibly be used as an aid in distinguishing between various assumed interaction potentials.

3.3. Force Constants and Vibration Spectra

The vibrational spectrum of a crystal is usually obtained from the Born-von Karman model of the crystal through the use of prescribed force constants (Born and Huang, 1954R). This model may be made completely general, but the usual scheme has been to consider force constants between first and second nearest-neighbor atoms only. Recent calculations have employed more general assumptions and have included force constants corresponding to ninth-nearest neighbors. In principle, these force constants may be obtained from calculations of the cohesion of the distorted crystal. This is a formidable problem which has so far not been completely solved, but developments in technique may bring it within the range of solution for simple metals in the near future. It is already evident from available experimental studies of vibrational spectra that it is necessary to include rather long-range interactions in the Born-von Karman model. The situation is particularly interesting when the Fermi energy makes a significant contribution to the elastic constants. The manner in which the energy of very short wavelength lattice distortions is determined then becomes of considerable importance since the Fermi energy is essentially a property of the infinite crystal. Theoretical investigations of this problem should throw considerable light on the meaning and generality of the band theory and should also contribute to the understanding of the interaction between electrons and atomic imperfections, since a short wavelength lattice distortion is in many ways similar to such an imperfection.

The recently developed experimental technique for obtaining vibrational spectra from the inelastic scattering of slow neutrons (see section 6) should stimulate further theoretical investigation of this problem. It is possible, in principle, to invert the observed spectrum at least formally to obtain all the microscopic elastic constants and hence, ultimately, the individual energy terms in the cohesive energy.

4. QUANTUM BASIS OF CRYSTAL CHEMISTRY

The development of a theoretical model for the calculation of the cohesive properties of intermediate crystal phases in binary (or multi-component) systems is hampered for several reasons. Serious limitations are introduced because of nonperiodicity of the lattice potential, complicated atomic structure, or combinations of

metallic, covalent, and ionic binding which characterize many of these materials. In solids of this type, ideas concerning phase stability are strongly guided by semi-empirical rules which are usually rationalized on the basis of a simplified model instead of being developed theoretically from first principles. As mentioned earlier, the often surprising simplicity of these rules offers some hope that they can be given a more fundamental theoretical justification which would, at the same time, provide an insight into the precise limits of validity of the rules. Many of these rules derive from simple theoretical notions which were introduced at the earliest stage of the development of band and valence theory; at the time, these rules were thought to have a justification or significance which later and more sophisticated understanding of the theory has shown to be invalid. Most of the effort during the past twenty years has been directed toward finding more experimental illustrations of the rules rather than in refining or justifying the theory.

4.1. Hume-Rothery Rules and Related Ideas

The most important Hume-Rothery (1936) rule from the metallurgical standpoint is that phases with metallic binding tend to exist at definite electron/atom ratios. This principle was first enunciated in 1926 by Hume-Rothery, who noted the constant electron/atom ratio at which each of the structurally analogous phases CuZn , Cu_3Al , and Cu_5Sn appear. The theoretical basis for phase stability in these so-called electron compounds was given by Jones (Mott and Jones, 1936) in terms of the earliest concepts of band theory using the approximation of nearly-free electrons. The phase stability is regarded as due to the lowering of the energies of some of the electrons relative to the free-electron values as the Fermi surface approaches a Brillouin zone boundary. This phenomenon occurs because the wavelengths associated with certain electrons become commensurate with the lattice dimensions of the crystal, thus giving rise to standing wave systems which can lower their potential energy in the lattice without significantly modifying their kinetic energy.

When described in this manner, a better appreciation can be gained that the conclusion might be independent of the specific model. Recent developments in many-body theory have shown that the Fermi surface has a real meaning, and that although individual particles have no real existence, there are excitations in the vicinity of the Fermi surface which not only have the character of quasi-particles but also have a characteristic wavelength. Stand-

ing wave effects can occur with these quasi-particles, suggesting that the explanation of the Hume-Rothery rules given by Jones may not be as specific to the free-electron model as it might appear superficially. The disappearance of these phases at higher electron/atom ratios is ascribed to the decrease in the density of electronic energy levels which occurs after the Fermi surface touches the zone boundary. Again, the argument depends primarily on zone geometry and may thus be retained in more rigorous theories.

This particular Hume-Rothery rule is most closely obeyed when the two alloying elements have nearly the same size and come from nearby positions in the periodic table. The rule apparently does not apply when the free-atom electronic energy levels of the pure constituent metals are widely different.

A second rule which connects phase stability with atomic size is as follows: It has been found experimentally that if the atomic radii of the constituent atoms in binary substitutional alloys differ by more than about 14 percent, or if the constituents come from widely separated positions in the periodic table, they do not form extensive solid solutions; instead, if they combine at all, they form ordered phases having a restricted composition range. The Hume-Rothery rules apply even when the interaction between ion cores is small, i.e., the size which is significant in the application of the rules is taken to be the interatomic distance in the crystal structure of the pure metal rather than the size of the ion core. Crudely, this behavior may be considered from the standpoint of the Fermi-Thomas approximation in the following way: the Fermi level may be thought to lie in approximately the same position in pure metals which are chemically similar. Thus, if metal atom A which is appreciably different in size is substituted into a matrix of metal B, its "Fermi level" will be severely displaced unless it is allowed to occupy approximately the same volume it does in the pure state. This requires a large relaxation of the B-lattice in order to accommodate atom A, and the resulting strain energy will suffice to inhibit the formation of the alloy phase.

There are also important size rules governing the stability of binary interstitial compounds. One of these, the Hägg rule, states that the maximum allowable expansion of the primary lattice resulting from the insertion of smaller atoms in interstitial positions is about 15 percent. A second rule is that the smaller atoms will occupy interstitial sites only when the relative atomic dimensions permit contact between the smaller atom and its neighbors. These empirical rules appear to be rather general and apply to both metallic and ionic systems. In metallic systems, the argument

stated in the preceding paragraph would appear to be relevant, but since most metallic interstitial phases involve the transition metals, a satisfactory theoretical discussion is complicated by the large number of valence electrons in the system and by the fact that these electrons are far from free. In ionic systems, it might be expected that the rules could be placed on a firm theoretical basis, but difficulties are experienced even here. Studies which have been made show that the difference in free energy between competing crystal structures is usually comparable with the theoretical uncertainty of the cohesive energy calculations for individual phases (Seitz, 1940).

Nevertheless, it might be desirable to calculate energy differences between competing structures for a homologous series of ionic compounds in which the ionic radius ratio varies systematically. Another suggestion is to study a particular elastic constant corresponding to a strain which is related to the phase transformation between the competing structures, e.g., c_{11} — c_{12} in the b.c.c.—f.c.c. transformation.

4.2. Coordination Number and Microscopic Symmetry in Crystals

Why an atom chooses a particular crystal structure, and why a few crystal structures are so prevalent in nature are questions which have no simple answer in terms of band theory. On the other hand, the valence bond picture, which specifically treats the interaction of an atom with its neighbors, has made considerable progress in this area (Pauling, 1960). For example, the 8-N rule for predicting the coordination number in covalent structures is explained on the basis of the number of bonding orbitals which are available. Furthermore, the frequent occurrence of the NaCl-type structure seems to have some connection with the preference of p orbitals for forming 90° bonds, while the existence of the competing CsCl-type structure is presumably associated with the bonds formed by sp^3 hybridization.

The valence-bond picture has been developed and extended considerably beyond its original scope, but this has led to an increasingly tenuous connection with basic theory. Pauling has extended the valence-bond description to metals by introducing the concept of a resonating system of bonds or, equivalently, fractional bonds. This approach has been followed by many investigators and has been applied to metallic and semiconducting phases as well as to complicated molecular systems. Pauling's ideas are by no means universally accepted, but at present they seem to provide the only available approach in many situations.

The principal difficulty in attempting to put the valence bond picture on a firmer basis is that it represents a perturbation approach which makes use of nonorthogonal wave functions. Furthermore, in the usual descriptions given by valence bond theory it is often difficult to determine whether molecular orbitals or highly correlated atomic orbitals are being considered. The requirement of orthogonal wave functions destroys the simple picture of the valence bond and leads to highly complex secular equations which are difficult to relate to the physics of the problem. Some progress has been made through the use of so-called valence bond functions, which are localized functions of the two-electron rather than the one-electron variety.

A number of interesting studies have been carried out which concern the change in coordination number within a particular column of the periodic table. Thus, in Group IVb, sp^3 hybridization produces strong tetrahedral bonds for the lighter elements, but the directional properties are considerably weakened by the increased number of radial nodes that are encountered in the heavier elements. At principal quantum number 5 (tin), the directional properties of the hybrids are strongly dependent on atomic radius, and there is a tendency toward higher coordination number. Thus, the high-temperature modification of tin (white tin) shows some vestiges of tetrahedral bonding, but the bond angles are warped, and two additional neighbors have moved into position. In lead, the transition to high coordination number is complete so that the structure is cubic. An analogous study of the change in structure in the series: S, Se, Te, Po, has been made by Von Hippel (1948). Systematic studies of this type, though qualitative, are extremely valuable in providing guide lines and suggestive ideas for the development of a more quantitative theory.

4.3. Relative Stability of FCC, BCC and HCP Structures in Metals

The relative stability of a particular crystal structure depends on small energy differences and thus requires a theoretical calculation of high accuracy. Even for the alkali metals in which cohesion is best understood, we are not certain why the body-centered cubic structure is preferred.

Zener (1947) has suggested that the body-centered cubic structure of the alkali metals is in reality a high-temperature phase, and following this suggestion, Barrett (1947) found a face-centered cubic modification of lithium at low temperatures; a close-packed modification of sodium has also been observed under comparable conditions (Barrett, 1956). Moreover, it is well known

that the body-centered cubic structure often appears as a high-temperature phase in other systems. Experimentally, body-centered cubic structures are usually found to have a higher specific heat and hence a higher entropy than the close-packed structures, but no detailed calculations of phonon entropy have yet been made.

Changes in structure in the transition metal series present another problem of considerable interest. Hopefully, this problem can be studied within the framework of band theory, but unfortunately the electronic wave functions are much more localized than the wave functions of the valence electrons in simple metals; as a result, many-body or correlation effects may be much more important. The problem has been treated from the valence bond standpoint, and the structural changes are explained by strong *spd* or *sd* hybridization. Some information about the mixed role of covalent and metallic binding in transition elements may possibly be obtained by studying the structures and stabilities of intermediate phases, i.e., phases such as the sigma phase, those having the α - and β -Mn structures, Laves phases of the AB₂ type, and various AB-type phases. Frequently, however, even the classification of such phases is uncertain, and there is a need for orienting studies, such as measurements of the low temperature specific heat and the temperature-dependence of the electrical resistivity, in order to ensure that the correct classification has been made.

4.4. Special Examples

4.4.1. Mooser-Pearson Semiconductor Rules

Mooser and Pearson (1956) have proposed a set of rules which are designed to govern semiconducting behavior in intermetallic compounds. These rules have been developed from the valence bond point of view, and appear to have no direct connection with band theory. On the other hand, they have been useful in predicting new semiconductor compounds, or new semimetals.

The first rule is an extension of the 8-N rule. It states that a necessary condition for semiconductivity is that

$$n_e/n_a + b = 8,$$

where n_e is the number of electrons per molecule, n_a is the number of Group IV to VII atoms per molecule, and b is the number of bonds between one of these atoms and other Group IV to VII atoms. Actually, the consequences of this rule also follow from band theory which states that a necessary condition for semiconductivity is an even number of electrons per unit cell of the crystal structure. Thus, in CdSb, the classic example discussed by Mooser

and Pearson, there are seven electrons per molecule; in order for semiconductivity to be possible within the framework of band theory, the unit cell must contain an even number of CdSb units. But from the Mooser-Pearson viewpoint a Cd-Cd bond must be present and this, of course, ensures an even number of cadmium atoms per unit cell.

The second Mooser-Pearson requirement for semiconductivity, namely, that all of the Group IV to VII atoms in the structure must be completely linked by covalent bonds, whereas a metal atom in the structure may have empty orbitals provided it does not have metal atoms as neighbors, has no direct analogy in band theory. But the rule is in accord with the general principle that semiconductivity requires at least partially covalent binding; pure ionic binding produces in most cases energy gaps which are too large, while the high coordination characteristic of metallic binding produces overlapping bands.

Attempts to refine the valence bond theory to explain such properties as electron and hole mobility, etc., are of doubtful validity since the model used ignores knowledge of the band structure determined experimentally.

4.4.2. The Refractory Hard Metals

The carbides, borides and nitrides of the three transition metal series are usually referred to as refractory hard metals (Schwarzkopf and Kieffer, 1953). These materials are metallic in character, often showing higher electrical conductivity than the parent transition metal. They have high melting points and low effective work functions and hence are of considerable technical importance; yet, very little work of a truly fundamental nature has been carried out with these materials.

The refractory hard metals have relatively simple crystal structures, the various phases following the rules for interstitial compounds discussed in section 4.1. Relatively little is known about their electronic structures although several hypotheses have been advanced. Umanski (1943) has proposed that these materials have the same band structure as the appropriate allotropic modification of the parent transition metal. This cannot be wholly true, however, because the nonmetal atoms contribute substantially to cohesion. Rundle (1948), working within the framework of the valence bond picture, assumes that covalent bonds are formed between the metal and the nonmetal atoms and that the directional property of the bonds accounts for the high cohesion and hardness of the materials. On the other hand, the model proposed by Rundle

requires a resonating system of bonds, similar to Pauling's metallic bond, to explain the metallic behavior of the refractory hard metals.

Apparently, the refractory hard metals as a class are not clearly describable by either the covalent bond or the usual band picture. It may therefore be particularly interesting to employ these materials in fundamental experiments designed to explore the band structure, such as low-temperature specific heat measurements, studies of the transport properties and elastic constants of single crystals, and even more detailed studies, e.g., the de Haas-van Alphen effect.

5. THEORY OF THE COHESIVE PROPERTIES OF ALLOYS

5.1. Fundamental Methods

The aim of a theory of metallic solutions must be primarily to explain the magnitude of the energy of solution of the solute atom in the solvent matrix. At the present time, however, a satisfactory method for computing solution energies is still lacking. The reason for this is that the solution energy, which is the difference between the energy of the solution and the energies of the separated pure components, is the small difference between two large quantities, both of which are subject to computational errors comparable in magnitude to the solution energy itself. The principal source of these errors arises from our inability to handle the many-electron problem. The best methods available are similar in main outline to those employed for pure metals and consist in using a one-electron approximation, viz., some form of the Hartree-Fock method, with added corrections designed to account for the effects of electron-electron correlation. We discuss below some of the specific problems connected with the computation as well as an evaluation of the methods currently available for handling these problems. Only dilute solutions will be considered here, a dilute solution being defined as one in which interactions between solute atoms can be neglected.

5.1.1. Screening

Unscreened Coulomb fields cannot exist at large distances from an impurity atom in the solvent lattice. This requirement of macroscopic electrical neutrality is probably satisfied by the many perturbation methods available. When screening is evaluated by present one-electron methods, a persistent disturbance in the charge density appears; this disturbance oscillates as a function of

distance from the solute with an amplitude which decreases as the inverse cube of the distance from the solute atom. In dilute solutions, the presence of these long-range oscillations is probably unimportant so far as the calculation of the solution energy is concerned. Satisfaction of the requirements for self-consistency at distances which are small compared to the Fermi wavelength is a different matter, however, and must not only be determined by more careful methods (especially where there is the possibility of a bound state) but is influenced in some measure by the considerations outlined in the next section.

5.1.2. Exclusion Principle and Correlation

When the solute atom introduces one or more extra valence electrons or reduces the number already present, special care must be taken to satisfy the exclusion principle and to account for electron-electron correlation. Even when no change in the number of conduction electrons is involved, correlation effects may be important. It is this complication that is mainly responsible for the difficulties in the calculation and for most of the uncertainty in the results. Virtually all calculations to date have employed the Hartree-Fock approximation, and hence a successful calculation of a solution energy has still not been obtained. It may be that some of the recent advances in the theory of the many-body problem will be applicable in some form to the calculation of solution energies, but such a conclusion has not yet emerged from current work.

5.1.3. Electronic Contribution to Binding

The determination of the electronic contribution to the binding of an impurity in a solvent matrix is, in principle, the simplest feature of the entire problem. This contribution is the sum of two terms: the mutual electrostatic interaction between the electrons and the ion cores, and the accompanying changes in the kinetic energy of the electrons. At present, this problem is usually treated by means of the Hartree-Fock approximation. The steps in such a treatment are:

- a. Introduce the perturbing atom into the lattice.
- b. Compute the modified conduction electron wave functions in the field of the perturbing atom and the other conduction electrons, including the effects of exchange with the perturbing atom core.
- c. Compute the resultant energy from the change of the Hartree-Fock energy parameters of the electrons and their mutual Coulomb and exchange interactions.

It may be noted that electron correlation effects are of importance in the more exact treatment of this problem, just as in the evaluation of Coulomb screening. Of some significance in this connection is the choice of the ion-core potential, a problem which is not much different in principle from that which arises in dealing with the pure crystal.

5.1.4. Ion-Core Repulsion and Lattice Relaxation

Except for the transition elements, it does not appear that any unusual complications arise in treating the problem of ion-ion interaction. These interactions can probably be dealt with adequately by methods already well-developed for molecular and ionic crystals. The simplest generalization that seems possible here is that, if the ion core is "small enough," the solute will have an interstitial state of not too high an energy. It is important to recognize that the long-range elastic strains in the lattice are simply relaxation strains due to the disparity between the size of the solute atom and the space into which it fits. The elastic strain energy contribution should not be included unless the change in the ion-core repulsion energy is also being carefully evaluated, since these contributions are not independent.

5.1.5. Critique

As a general critique on the methods available for calculating solution energies, the following items deserve special mention:

- a. The best attempts at a first-principle calculation which have been made so far involve methods that are committed to computing the energy of solution as the difference between two large quantities. It is important to develop methods for formulating the problem as a true perturbation problem so that errors associated with calculations of the absolute energies of the pure components and the dilute alloy can be circumvented.
- b. It is probable that the difficult part of the problem, viz., that of correlation and self-consistency, can be handled more readily by restricting attention to the fictitious problem of the solution of an atom in a "free" electron gas. The complications introduced by the crystal lattice are not fundamental to the treatment of this part of the problem.
- c. It is desirable that other simple equilibrium properties of the solute state also be well represented in the course of the calculation. Of particular concern here are the spin and orbital magnetic properties.

5.2. Models and Semi-Empirical Methods

A number of models or semi-empirical approaches have been developed which seek to relate the properties of the solution phase to certain characteristics of the component substances without rigorous justification of the assumptions involved. By introducing simple models having undetermined or adjustable parameters, an attempt is made to rationalize the systematic variation in properties which is frequently observed, for example, within a homologous series of intermetallic phases. These models are valuable as a means for correlating experimental data and as conceptual frameworks within which new semi-empirical relationships may be sought. Quite commonly, one or at most a few significant fundamental parameters of the system are varied in a model in which all of the other properties are known or predetermined. Treatments of this type are useful so long as the model is not made too complicated or too specific. Unfortunately, much of the published work in this area suffers from over-refinement of the models employed; the models thus acquire a strong *ad hoc* flavor.

The following models are fairly representative of the various approaches which have been found helpful, not only in suggesting experimental programs but also in providing insight into some of the properties of alloys.

5.2.1. The Rigid-Band Model

The model characterized as the rigid-band model was proposed by Jones and has already been discussed briefly in connection with the Hume-Rothery rules in section 4.1. The elements of the model are a density-of-states function as determined from the free-electron approximation, and the idea that each solute atom contributes all of its valence electrons to the conduction band so determined. The density-of-states function depends principally, of course, on the crystal structure and on the unit cell dimensions.

It is found by first-order perturbation theory, using plane wave, zero-order wave functions, that the presence of the solute ion produces a constant shift in all electronic energy levels. Therefore, in this approximation, the density-of-states function is unchanged except for a uniform shift in the energy of the band as a whole. The properties of the alloy are then deduced from the new electron distribution that results from filling the band to the level corresponding to the number of electrons per atom in the alloy.

This model was remarkably successful in rationalizing some of the empirical rules formulated by Hume-Rothery, and it has been the prototype of many subsequent models. Despite its early useful-

ness, however, attempts to elaborate or extend the model to systems and properties other than those originally considered have been generally unsuccessful.

5.2.2. The Elastic Strain Model

It is generally accepted that the size difference between solute and solvent atoms is an important consideration in the fundamental theory of solution energetics. On the other hand, attempts have been made to explain the energy of solution solely in terms of a quasi-continuum model which involves certain elastic properties and perhaps a few additional parameters. Since the accommodation strains in the matrix caused by the insertion of a solute atom always produce a positive energy contribution, it is necessary to introduce rather arbitrarily into the model some relatively negative contribution, such as a surface or interfacial energy contribution.

Such models are generally not very successful. They suffer principally from the assumption that elastic strain energy considerations play a dominant role in the solution process. The strongest criticism that can be leveled against such models is that they predict drastic changes in the heat of solution at the melting point because the liquid metal no longer has any static rigidity, whereas the observed differences are considerably smaller.

5.2.3. Statistical Models

Attempts to explain the cohesive properties of solids and alloys on the basis of a statistical model of the Thomas-Fermi type have not enjoyed much favor or support. Although the literature pertaining to this subject is fairly extensive, much of it has been contributed by a comparatively few individuals, and the general approach does not constitute a popular part of the current endeavor. Nevertheless, it seems likely that, with past and future improvements in the formalism, this type of treatment may hold the best promise for the development of a useful, semiquantitative model for the solution process and possibly even for the fundamental features of the cohesion of metals.

5.2.4. Critique

On the whole, semi-empirical models for the cohesive properties of solutions have not proved very successful, although they are of considerable heuristic value as noted above. Some danger exists in drawing conclusions about the validity of specific models on the basis of limited quantitative agreement. The best protection

against error of this sort is provided by seeking model-independent bases for relationships that may have been suggested originally by specific models.

It is worth noting, furthermore, that there are a number of equilibrium properties which can in principle be represented by models of the same type as those used to describe the simple transport properties. Some of these equilibrium properties are electronic paramagnetism and nuclear magnetic resonance. They share the distinction of being one-electron properties which depend on continuum excitation. It can be expected that these properties might be adequately described, along with the transport properties, by relatively simple solution models that deal only with the modification of the electron wave functions and density-of-states function at the Fermi level.

There is considerable room for improvement in the construction of models for representing the cohesive properties and other equilibrium properties of alloy phases. It might be expected that models of this type would also provide a more precise basis for representing nonequilibrium properties or for testing corresponding models already developed for this purpose.

6. METHODS OF MEASUREMENT OF COHESIVE PROPERTIES ¹

6.1. Cohesive Properties of Pure Crystals

A knowledge of the cohesive properties of a crystal, i.e., the binding energy and the microscopic elastic constants, would in principle ² allow the calculation of the thermodynamic functions of the crystal, and hence of its equilibrium properties. At the present time it is more realistic to think of the calculation in reverse, i.e., to consider the use of experimentally determined properties to evaluate the binding energy and elastic constants.

6.1.1. The Cohesive Energy

The experimental quantity most closely related to the computed cohesive energy is the heat of sublimation of the crystal. The comparison that is actually made, however, depends on the nature of the crystal, inasmuch as this may affect both the choice of the reference state in the theoretical computation and the kind of experimental measurements from which the heat of sublimation is derived.

¹ For general references see Kubaschewski and Evans (1958R) or Aston and Fritz (1959R).

² In the harmonic approximation.

a. The theoretical reference state. The theoretical cohesive energy is hardly ever computed with respect to the completely ionized nuclei and electrons. For metals and covalent crystals, the usual procedure is to separate the atoms into "cores" and "valence electrons," and thus to compute the energy of the crystal with respect to the corresponding ions and electrons at rest at infinity. In addition to thermodynamic data, a need therefore exists for spectroscopic data on the ionization energies of the neutral atoms.

b. The experimental determination of cohesive energies. It is not usually possible to measure the cohesive energy of a metallic, covalent, or ionic crystal directly as the work needed to dissociate it completely into its constituent atoms, ions, or molecules. Instead, the vapor pressure of the crystal (or the liquid) is measured as a function of temperature, and the thermodynamic relation (Clausius-Clapeyron equation)

$$\frac{dp}{dT} = \frac{\Delta H}{T\Delta V} \quad (6.1)$$

is used to calculate ΔH , the enthalpy of vaporization per gram atom. Here p is the vapor pressure and ΔV is the volume increase on vaporizing one gram atom of the condensed phase. If the vapor is ideal, ΔV is just RT/p less the molar volume of the condensed phase, V_m , and the energy of vaporization, ΔE , differs from ΔH only by $RT - pV_m$. In general, particularly at high pressures, it is necessary that the equation of state of the vapor be known in order to obtain ΔV at the temperature in question.

In the simplest case, namely, when the vapor is composed of the atoms in the theoretical reference state, the desired cohesive energy is obtained by extrapolating ΔE so measured to 0°K , and adding to this value the zero-point vibrational energy of the crystal. For the vapor, this requires a knowledge of the heat capacity, which can usually be calculated very accurately from spectroscopic data. For the condensed phase, heat capacity data are needed from low temperatures to the temperatures at which the vapor pressure is measured, and if these temperatures are above the melting point, the heat of fusion is required as well.

A common complication is the presence in the vapor of appreciable concentrations of diatomic and polyatomic molecules. This state of affairs is nearly always encountered with covalent and ionic solids, and polymerization in the vapor phase is often observed with metals, too. Methods must then be contrived for measuring or calculating the vapor composition, and for determining the molecular dissociation energies. The standard methods of

statistical mechanics can be used if sufficient molecular data can be obtained, spectroscopically or otherwise. Absorption spectra of the vapor may be used for analytical purposes. Mass spectrometric analyses of the vapor are becoming increasingly useful, especially for the measurement of high-temperature equilibria. Obviously, the method just described fails if the crystal in question has such a low volatility that the equilibrium vapor pressure is below the limit of measurement. It may then be possible to obtain the cohesive energy directly from calorimetric measurements of the energy of formation of the crystals from components whose vapor equilibria are measurable or calculable.

The energy difference between two allotropic modifications can be determined calorimetrically if both forms exist at atmospheric pressure; it is convenient, but not necessary, if equilibrium between the two forms is reached rather rapidly. Allotropic forms which exist only at high pressures are at present much less accessible to direct measurement of their cohesive energies. However, the Clausius-Clapeyron Eq. (6.1) can also be used in such situations provided equilibrium can be achieved with a phase of known energy. The equilibrium pressure must be found as a function of temperature, and the volume change accompanying the transition must be determined, either dilatometrically or from X-ray diffraction data for the two crystal forms.

The cohesive energy of a metastable modification can be obtained in principle by extrapolation of thermodynamic data for solid solutions having the same crystal structure; this procedure cannot at present lead to trustworthy results, partly because adequate theories of solid solutions do not yet exist, and partly because of the experimental uncertainties.

6.1.2. Microscopic Elastic Constants³

The elastic constants needed to specify the potential energy of a crystal as a function of the nuclear coordinates are far too numerous to be determinable from the macroscopic elastic properties alone. However, they can be evaluated from the vibrational spectrum of the crystal, which is accessible in several ways:

a. Spectroscopy in the conventional sense, i.e., absorption and inelastic scattering (Raman effect) of light, has so far yielded scant information about the complex vibrational spectra of crystals. Only the highest frequency vibrations of some ionic and covalent crystals can be excited optically, and the resolution of

³ The measurement of macroscopic elastic constants is described by Huntington (1958R).

Raman effect measurements is not nearly adequate to reveal any of the fine-scale detail within the essentially continuous spectrum. The strong continuous electronic absorption in metals, of course, precludes optical studies of vibrational spectra.

b. Inversion of specific heat data has been proposed several times as a method for determining the vibrational spectra of solids. The insensitivity of such thermodynamic functions to the detailed energy level distribution is a severe limitation on this method, in spite of the high experimental accuracy attainable. However, some progress has recently been made in this direction in favorable cases (Barron, Berg and Morrison, 1957).

c. Diffuse scattering of X-rays. The non-Bragg scattering of monochromatic X-rays by perfect crystals consists of incoherent (Compton) scattering and coherent scattering by elastic waves. If the contribution due to incoherent scattering can be calculated or measured independently, the coherent scattering in special directions can be used to evaluate dispersion curves for purely longitudinal or purely transverse waves in these directions. The corresponding microscopic elastic constants can then be calculated with the aid of the Born-von Karman theory. The accuracy obtainable at present is not very great, and the measurements are difficult, but as techniques are improved this situation may be remedied to some extent.

d. Inelastic scattering of slow neutrons. Slow neutrons may be inelastically scattered by solids, with the creation or destruction of a phonon (quantum of elastic vibration). In the limiting case, where the neutron is initially at rest, the energy and momentum of the scattered neutron are just those of the phonon which was destroyed. It is possible to produce neutron beams from cold scatterers (ideally, either liquid or solid hydrogen) whose mean energies are low compared to typical maximum vibrational energies, and to measure the energy distribution of these neutrons after Bragg reflection from the crystal. Thus, dispersion curves may be constructed, and again these may be interpreted in terms of the Born-von Karman theory. Several such studies have already been carried out (e.g., Brockhouse and Iyengar, 1958), and the method appears to be very promising.

6.1.3. Present Status of Knowledge of Cohesive Properties

The cohesive energies of a large number of simple solids—metallic, covalent, ionic, and molecular—are known with accuracies which are, for the most part, far in excess of those needed to test present-day calculations. The methods described in section 6.1.1

are well developed and are gradually being improved and extended to a wider variety of materials, so that the experimental coverage is reasonably satisfactory. Nevertheless, further studies and measurements of this kind would be highly desirable.

Our present knowledge of the vibrational spectra of solids is, by contrast, very meager. This situation is in part due to the newness of the techniques and to experimental difficulties, but also, and especially for the neutron scattering technique, to the very great investment in apparatus that is necessary.

6.2. Cohesive Properties of Alloys and Other Multicomponent Systems

6.2.1. The Cohesive Energy

The binding energy of a solid solution or other intermediate phase relative to the energies of the separated atoms is both more difficult to measure and of less immediate interest than the binding energy relative to the pure components. Furthermore, most alloy theories aim to describe the changes in properties relative to one or both components, rather than to treat the alloy as an independent system.

Direct measurements of heats of formation of solid alloys from the pure components are rarely possible, because diffusion is generally so slow that exceedingly long times are required to establish equilibrium. However, the heat of formation can frequently be determined as the difference between the heats of solution of the pure components and of the alloy itself in a common solvent. The solvent is chosen, other things being equal, to minimize the individual heats of solution; thus, liquid metals (e.g., tin) are usually the best solvents for alloys.

From measurements of the thermodynamic activities of the components of an alloy, the free energy of formation of the alloy can be calculated. In a binary system, the activity of only one component need be measured, since the activity of the other component may be obtained by using the Gibbs-Duhem relation. Nevertheless, measurement of the activities of both components is preferable, though rarely possible. Activities can be determined from the electromotive force of concentration cells if the alloy forms a reversible electrode in contact with a suitable electrolyte. When one component of the alloy is sufficiently volatile, its vapor pressure, and hence its activity, can be measured by standard methods. Radioactive tracer and mass spectrometric techniques are often of special value in this connection. The accurate determination of phase boundaries is useful, mostly as a check on activi-

ties determined by other methods. The temperature coefficient of the thermodynamic activity at fixed composition gives the relative partial molal entropy of a given component, and this in conjunction with the free energy of formation, can be used to compute the heat of formation of an alloy. This procedure can be employed when a direct determination is difficult, but it is generally less reliable than a direct measurement.

All of the methods just described give the enthalpy of formation of an alloy at the temperature of measurement. If the energy of formation in the ground state is needed, it must be obtained by extrapolation of equilibrium enthalpy values to 0°K. This procedure is much more hazardous for alloys than for pure metals because the equilibrium atomic distribution is difficult if not impossible to achieve at low temperatures, where diffusion takes place very slowly. Therefore, heat capacity measurements below some temperature (say half the melting point) will include only changes in vibrational energy, and configurational contributions will be missed because they are "frozen in." Only in systems in which long-range order exists is there any possibility of attaining the ground state with respect to configurational energy as well as vibrational energy.

6.2.2. Microscopic Elastic Constants

The methods described in 6.1.2 for the determination of vibrational spectra are also applicable to alloys. However, possible contributions to the specific heat arising from changes in configurational energy represent an additional limitation on Method b, and in disordered alloy systems a further contribution to the diffuse X-ray scattering is present which affects the use of Method c.

6.2.3. Present Status of Knowledge of Cohesive Properties of Alloys

The relative thermodynamic functions of a large number of solid solutions, mostly of metals, are known with varying accuracy. That this accuracy is often much greater than can be attained theoretically is more a reflection on the inadequacy of the theoretical situation than on the excellence of the experimental work. If refinements were soon to appear in phenomenological theories of solid solutions, there might prove to be an embarrassing shortage of really accurate data with which to test such new developments. A need probably also exists for accurate thermodynamic data on a wider variety of systems, since data of this kind may suggest or help to formulate empirical generalizations.

We are almost completely ignorant of the vibrational spectra of solid solutions. The low temperature heat capacities of a few alloys have been measured with sufficient accuracy to show deviations from additivity, but in no case have the relatively accurate methods described in section 6.1.2. been applied to any alloy system.

7. REFERENCES

7.1. Reviews and General Articles

- Aston, J. G. and Fritz, J. J., 1959, *Thermodynamics and Statistical Mechanics* (New York, John Wiley and Sons, Inc.).
- Born, M. and Huang, K., 1954, *Dynamical Theory of Crystal Lattices* (Oxford, Oxford University Press).
- Brooks, H., 1958, *Nuovo Cimento*, 7(Suppl.), 165.
- Callaway, J., 1958, "Electron Energy Bands in Solids," in *Solid State Physics* (New York, Academic Press), Vol. 7, p. 99.
- Cohen, M. H. and Heine, V., 1958, *Phil. Mag.* 7(Suppl.), 395.
- Harrison, W. A. and Webb, M. B., Eds., 1960 *The Fermi Surface* (New York, John Wiley and Sons, Inc.).
- Hume-Rothery, W., 1936, *The Structure of Metals and Alloys* (London, Institute of Metals).
- Huntington, H. B., 1958, "The Elastic Constants of Crystals," in *Solid State Physics* (New York, Academic Press), Vol. 7, p. 213.
- Kubaschewski, O. and Evans, E. L., 1958, *Metallurgical Thermodynamics* (New York, Pergamon Press).
- Löwdin, P. O., 1956, *Phil. Mag.* 5(Suppl.), 1.
- Mott, N. F. and Jones, H., 1936, *Theory of the Properties of Metals and Alloys* (Oxford, Clarendon Press).
- Pauling, L., 1960, 3d edition, *The Nature of the Chemical Bond* (Ithaca, New York, Cornell University Press).
- Pines, D., 1955, "Electron Interaction in Metals," in *Solid State Physics* (New York, Academic Press), Vol. 1, p. 368.
- Reitz, J., 1955, "Methods of the One-Electron Theory of Solids," in *Solid State Physics* (New York, Academic Press), Vol. 1, p. 2.
- Schwarzkopf, P. and Kieffer, R., 1953, *Refractory Hard Metals* (New York, MacMillan Co.), Chap. 3.
- Seitz, F., 1940, *Modern Theory of Solids* (New York, McGraw-Hill).
- Wigner, E. P. and Seitz, F., 1955, "Qualitative Analysis of the Cohesion in Metals," in *Solid State Physics* (New York, Academic Press), Vol. 1, p. 97.

7.2. Original Articles

- Allen, L. C. and Karo, M., 1960, *Revs. Modern Phys.*, 32, 275.
- Amdur, I., 1958, *J. Chem. Phys.* 28, 987.
- Barrett, C. S., 1947, *Phys. Rev.* 72, 245.
- Barrett, C. S., 1956, *Acta Cryst.* 2, 671.
- Barron, T. H. K., Berg, W. T., and Morrison, J. A., 1957, *Proc. Roy. Soc. (London)*, A242, 478.
- Brockhouse, B. N., and Iyengar, P. K., 1958, *Phys. Rev.* 111, 747.
- Brooks, H. and Ham, F. S., 1958, *Phys. Rev.* 112, 344.

- Daniels, W. B. and Smith, C. S., 1958, *Phys. Rev.* **111**, 718.
- Ham, F. S., 1955, "The Quantum Defect Method," in *Solid State Physics* (New York, Academic Press), Vol. 1, p. 127.
- Harrison, W. A., 1959, *Phys. Rev.* **116**, 555.
- Heine, V., 1957, *Proc. Roy. Soc. A240*, 340, 361.
- Hirschfelder, J. O., Curtiss, C. F. and Byrd, R. B., 1954, *Molecular Theory of Gases and Liquids*, (New York, John Wiley and Sons, Inc.) p. 179.
- Kip, A. F., 1960, "Cyclotron Resonance in Metals-Experimental," in *The Fermi Surface* (New York, John Wiley and Sons, Inc.) pp. 146-154.
- Kohn, W. and Rostoker, N., 1954, *Phys. Rev.* **94**, 1111.
- Leigh, R. S., 1951, *Phil. Mag.*, **42**, 876.
- Löwdin, P. O., 1948, Thesis, Uppsala.
- McIrvine, E. C. and Overhauser, A. W., 1959, *Phys. Rev.* **115**, 1531.
- Martin, P. C. and Schwinger, J. C., 1959, *Phys. Rev.* **115**, 1342.
- Mooser, E. and Pearson, W. B., 1956, *J. of Electronics* **1**, 629.
- Nozieres, P. and Pines, D., 1958, *Phys. Rev.* **109**, 741.
- Raimes, S., 1952, *Phil. Mag.* **43**, 327.
- Rundle, R. E., 1948, *Acta Cryst.*, **1**, 180.
- Seitz, F., 1935, *Phys. Rev.* **47**, 400.
- Seitz, F., 1940, *Modern Theory of Solids* (New York, McGraw-Hill Book Co., Inc.), p. 265.
- Shoenberg, D., 1960, "The de-Haas-van Alphen Effect," in *The Fermi Surface*, (New York, John Wiley and Sons, Inc.) pp. 74-84.
- Smith, Charles S., 1959, "Relation of Metallic Bonding to Elastic Constants," in *Resonance and Relaxation in Metals* (Cleveland, American Society for Metals).
- Stern, F., 1959, *Phys. Rev.* **116**, 1399.
- Umanski, J. S., 1943, *Ber. Akad. Wiss. USSR—Physik-chem. Analyze*, **16**, 127.
- Von Hippel, A., 1948, *J. Chem. Phys.* **16**, 372.
- Wigner, E. P. and Seitz, F., 1934, *Phys. Rev.* **46**, 509.
- Woodruff, T. O., 1957, "The Orthogonalized Plane-Wave Method," in *Solid State Physics* (New York, Academic Press), Vol. 4, pp. 367-411.
- Zener, C., 1947, *Phys. Rev.* **71**, 846.

8. APPENDIX: COHESION OF METALLIC ALLOYS— A SHORT THEORETICAL SURVEY

J. FRIEDEL

The following more detailed discussion of alloy theory based on specific models is included separately as an appendix in order to avoid breaking the general continuity of the preceding argument.

8.1. Main Factors Affecting the Cohesion of Alloys

The main factors have been analyzed in Hume-Rothery's (1931) classical work and have been reviewed more recently by Friedel (1954, 1958). They may be summarized as follows:

(a) A difference in electronegativity of the components of the alloy, producing localized charges on the various atoms, thus lead-

ing at certain concentrations to the formation of ionic compounds (e.g., NaCl). The fact that order is destroyed at high temperatures in phases such as β -CuZn shows that this term is small for alloys consisting of "ordinary" metals which are close to each other in the periodic table.

(b) A covalency factor which is important only for certain components when present in phases where homopolar bonds can be formed, e.g., InSb in the diamond-cubic structure (*sp* bonds) and transition metal alloys in the body-centered-cubic structure (*sd* bonds).

For components with *s* and *p* valence electrons ("ordinary" metals), these two factors stabilize "valency" compounds with an average number of valence electrons per atom equal to 8. These compounds are often fairly stable and are sometimes insulators or semiconductors.

(c) For ordinary metals, in which conditions do not favor the formation of valency compounds, and also for the transition metals, a "metallic binding" occurs, frequently accompanied by extensive primary solid solubilities. This type of binding is often interpreted as a sharing by all the atoms of their valence electrons, which behave essentially as a gas of nearly free electrons. Using such a picture, Jones (1934) was able to explain the occurrence of the "Hume-Rothery phases" at definite electron/atom ratios, corresponding to which the Fermi surface of the alloy presumably touches the limits of a Brillouin zone of the crystal structure. It should, however, be kept in mind that in these alloys the valence electrons are not uniformly shared between the atoms; instead, the valence electrons are undoubtedly not as free as assumed in the approximate treatment given by Jones. Moreover, in (partly) ionic or covalent alloys, electrons are also shared and they are only moderately less free than those in "ordinary" alloys. Finally, the explanation proposed by Jones, although valid for nearly free electrons, cannot be applied to *d* electrons of transition metal alloys (e.g., σ -phases).

(d) Transition metal, rare earth, or heavy metal atoms often have strong, localized magnetic moments in alloys. These no doubt influence the cohesive properties, but in a way which is still obscure because the magnetic interactions themselves are not well understood. The highly localized 4 *f* shells of the rare earth metals must probably be distinguished from the more extended *d* shells of the transition metals. The latter lead to *d* bands at high concentrations, and to *d* virtual bound states at low concentrations. The 5 *f*

shells in alloys of the heavy metals are probably intermediate in behavior to that of the covalent and localized types.

(e) Finally, solute atoms readily enter interstitial positions in a metal only if they are considerably smaller in size than the solvent atoms, and extensive solubility is obtained in substitutional alloys only when the solute atoms have roughly the same size as the host atoms. Solute atoms having intermediate sizes may tend to form compounds (e.g., Laves phases). This size factor and also the variation of the elastic constants with composition are usually treated in terms of classical elasticity, assuming that the components of the alloys have the same equilibrium sizes and elastic constants as the pure components.

8.2. Usual Approximations

Various approximations have been employed in considering the cohesive properties of metallic alloys. In this section the validity of these approximations will be tentatively assessed.

It is customary to distinguish the purely electronic factors—electronegativity, covalency, etc.—from the size factor, which involves, more directly, displacements of the nuclei. Even without taking the displacements of the nuclei into account, the study of alloys is basically difficult, because it involves many electrons ($\approx 10^{24}$), moving in a nuclear potential of low symmetry. The first difficulty is ignored in the independent electron model; the second is usually handled by perturbation methods starting from either the free-electron or tight-binding approximation for the pure metal.

8.2.1. Independent-Electron Model

In a pure metal, each electron is assumed to move in a fixed average periodic potential. This potential contains average corrections for correlation effects but is usually assumed to be the same for all the electrons, which move in extended states throughout the entire volume of material. The independent-electron model is often extended to alloys by assuming that all the electrons are still moving in the same average potential. This leads to the following difficulties:

(a) The average correlation corrections have never been properly computed for an alloy. These correlation corrections are very important for cohesion, because they affect the energy of each electron, and also because the average correlation energy must be subtracted from the sum of the one-electron energies to obtain the total energy of the system. Both these points become significant

when covalent or ionic factors are involved; even for metallic binding, they are important in computing the energy of solution and the interaction energies between solute atoms.

(b) The average potential is no longer periodic. The often strong perturbations which are introduced may lead to a local accumulation of electronic charge, appearing as bound states (e.g., $4s^2$ states in Cu-Ga alloys, $4s^2$, $3d^{10}$ states in Cu-Ag alloys) or virtual bound states (e.g., Cu-Ni, Cu-Co, and Cu-Fe alloys). It is clear that the correlation corrections for these states should not be the same as those for other extended states. However, if different average potentials are used for different states, these will no longer be orthogonal. This difficulty does not arise with alloys of the transition elements having permanent magnetic moments (e.g., Cu-Mn alloys) when each of the two spin directions is treated in its own average potential; the states are still orthogonal through their spin parts.

Only a few attempts have been made to estimate the correlation corrections in alloy systems. An average Thomas-Fermi type of correction has been used by Huntington and Seitz (1942), Friedel (1952) and others to compute self-consistent average potentials (for problems involving vacancies and interstitials in copper or "ordinary alloys" such as Cu-Zn, Cu-H, etc.). Friedel (1952) has sometimes assumed that when bound states appear (e.g., in Cu-Ga alloys) they have the same electronic structure as in the isolated ions. Varley (1954) has assumed that each atom has the same correlation energy as it would have in the corresponding pure element. In virtual bound states (e.g., in Cu-Mn or Al-Mn alloys) Friedel (1956) has assumed that the exchange terms are the same as in the free ion, except for quenching of the orbital angular momentum.

These estimates are probably of the right order of magnitude but have shaky theoretical foundations. Kohn et al. (Langer and Vosko, 1960) have recently computed a self-consistent potential, including correlation effects, for a free-electron gas surrounding a fixed nuclear charge, using methods of field theory. It would be of great interest to compute the correlation energy by similar methods for realistic electron densities.

8.2.2. Perturbation Methods

Perturbation methods were used almost exclusively in all theoretical work on alloys carried out prior to about 1940. These methods take complementary forms in energy and in space.

(a) *Energy.* Treating the change, V_p , in the lattice potential

due to alloying as a perturbation, the change in energy of a (normalized) state $\psi(E)$ of the pure matrix is

$$\Delta E = \int_V \psi^* V_p \psi d\tau + \sum' \frac{\left| \int_V \psi^* V_p \psi' d\tau \right|^2}{E - E'} + O_3(V_p), \quad (8.1)$$

where the summation extends over all the other states $\psi'(E')$ of the matrix.

The first term in ΔE usually varies smoothly with energy. The band structure of the alloy may then exhibit a density of states rather similar to that of the matrix, provided that: a) V_p is small; b) E is not too near a band edge (e.g., the Fermi level); and c) for some energies, E , V_p is not periodic or does not have a periodicity which is a multiple of that of the matrix (no order); the large contributions of the second term for $E' \cong E$ with $E' > E$ and $E' < E$ then cancel each other.

If, furthermore, a free-electron approximation can be used, i.e., if

$$\psi \exp(-ik \cdot r) \cong U_0, \quad (8.2)$$

independent of E , then

$$\Delta E \cong \int_V U_0^* V_p U_0 d\tau. \quad (8.3)$$

The band structure of the alloy is then deduced from that of the pure solvent by a mere shift, ΔE , independent of E ; this is the rigid-band approximation. This approximation is the basis of Jones' explanation of the Hume-Rothery phases and of Mott's (1935) treatment of the d -band in Ag-Pd or Cu-Ni alloys. Although recent work has confirmed the usefulness of this approximation, little in general can be said concerning its exact validity. The rigid-band approximation should of course be better for the valence band of "ordinary" alloys (e.g., Cu-Zn) than for the d -band of transition metal alloys (Ag-Pd, etc.).

The second term in Eq. (8.1) is usually not negligible, even for a small perturbation such as that introduced when Zn is substituted for Cu in Cu-Zn alloys. A study of this term shows that alloying should tend to broaden the bands, and thus to decrease the density of states. The neighborhood of a band edge has also been studied both by perturbation methods and by numerical computation for one-dimensional as well as finite lattices. The bound states which might appear for small concentrations and large perturbations tend to broaden into an exponential tail at high concentrations. The details of the density-of-states curve, $n(E)$, near the limit should also be somewhat blurred by the second-

order terms. This effect of a Brillouin zone boundary has not been studied so far.

It has recently been suggested by Cohen and Heine (1958) that, near the Fermi level of ordinary alloys, such as Cu-Zn, the free-electron approximation, Eq. (8.2), is not valid; the change of character of $u = \psi e^{-ik \cdot r}$ from s to p with increasing energy should lead to a variation of ΔE with energy and thus to a change in the form of the band. Only qualitative arguments and rather unconvincing experimental evidence for this effect have been given. It should nevertheless be considered.

(b) *Coulomb Interactions.* To compute correctly the Coulomb interactions, it is essential to complement the energy structure of the alloy with a picture in space. First-order approximations have been used in different ways, depending on whether a free-electron approximation is valid or not.

Long-range oscillations. The perturbation method gives new wave functions of the form

$$\Psi + \Delta\Psi = \Psi + \sum' \frac{\int_v \Psi'^* V_p \Psi d\tau}{E - E'} \Psi'. \quad (8.4)$$

Summing these contributions over all occupied states Ψ , it can easily be shown by dimensional reasoning that a change in the electronic density is introduced which decreases at large distances r from a solute atom as r^{-3} times an oscillating function of r with wave length of the order of the Fermi wave length (Bloembergen and Rowland, 1953). These long-range oscillations persist when correlation terms are taken into account (cf. Langer and Vosko, 1960). They are small outside the perturbed region. Neglecting these oscillations leads to the Thomas-Fermi approximation.

Linearized Thomas-Fermi type of approximation. If V_p were a constant in space, the density $\rho(E, r)$ of electrons in the alloy having energies less than E would be related to the corresponding density $\rho_0(E, r)$ in the pure solvent by

$$\rho(E, r) = \rho_0(E - V_p, r). \quad (8.5)$$

When V_p varies slowly with r , e.g., at large distances from a solute atom, Eq. (8.5) is still approximately valid and may be regarded as a generalized Thomas-Fermi approximation, which can be justified in the same way. Using Poisson's equation and treating V_p as a perturbation leads to the basic equation of the linearized Thomas-Fermi approximation:

$$\Delta V_p = -4\pi \left[\rho(E'_M, r) - \rho_o(E_M, r) \right] = 4\pi \frac{d\rho_o}{dE_M} (V_p + E_M - E'_M) + O_2(V_p + E_M - E'_M), \quad (8.6)$$

where E'_M and E_M are the Fermi levels of the alloy and the pure solvent.

The electronic density at the Fermi level of the solvent is $d\rho_o/dE_M$. This quantity has the periodicity of the lattice, and its average value in space is the density of states $n_o(E_M)$. Eq. (8.6) implies that each solute atom introduces a perturbation V_i which decreases exponentially at large distances r_i , and that these perturbations are additive; i.e.,

$$V_p = \sum_i V_i,$$

with

$$V_i = f(r) e^{-qr_i} \text{ for } qr_i \gg 1, \quad (8.7)$$

$$q^2 = 4\pi n_o(E_M),$$

and

$$E'_M = E_M.$$

In Eq. (8.7), $f(r)$ is a function having the periodicity of the lattice; $f(r)$ can be taken as a constant if a free-electron approximation is valid.

The Fermi level is unaltered by alloying, a quite general result for noninteracting solutes. Values of n_o for usual metals are such that each solute atom should only perturb its first-nearest neighbors. These perturbations should be small, especially when the solvent is a transition metal which has a high density of states, $n_o(E_M)$.

This result becomes self-evident when it is remembered that, within the limits set by classical physics, an electric field cannot exist within a metallic substance; differences in nuclear charges must be screened out within distances comparable with the Fermi wavelengths, i.e., within a few interatomic distances. This point, first emphasized by Mott (1936), is not in contradiction with the rigid-band picture. It can indeed be shown that in the free-electron approximation ($f(r) = \text{const.}$), the linearized Thomas-Fermi approximation in space leads to the rigid-band model of the energies.

Born approximation. When the valence electrons of the solvent can be regarded as free, their scattering by the lattice potential, V_p , can be treated by the standard first-order (Born)

approximation. Scattering is proportional to $V_p - \bar{V}_p$, where \bar{V}_p denotes a spatial average. This explains qualitatively why the electrical resistivity varies as $c(1 - c)$ in disordered solid solutions, c being the atomic concentration of the solute (Nordheim, 1931), or as the square of the valence difference, Z , between solute and solvent (Mott, 1936), and, in ordered alloys, roughly as $1 - S^2$, where S is the long-range order parameter (Muto, 1936). This approximation is strictly equivalent to a linearized Thomas-Fermi model. Despite its qualitative success in explaining electrical properties, it can only be justified for very small perturbations, such as those which exist in disordered β -CuZn alloys ($Z_{\text{eff}} = 1/2$). For $Z > 1$ (copper-base alloys containing Ga, Ge or As), the perturbations are certainly too large to be treated by a simple Born approximation.

8.2.3. Free-Electron Approximations

The valence electrons of an "ordinary" solvent are often treated as a gas of free electrons scattered by the solute atoms. It must be pointed out that free-electron approximations of this type are not valid for d electrons in transition metal alloys or sp electrons in covalent structures, both of which should be treated by other methods (tight binding). Free-electron approximations are also not satisfactory for studying Brillouin zone boundary effects.

For a pure metal A, the Schrödinger equation is

$$\frac{\hbar^2}{2m} \Delta \Psi + (E - V_A) \Psi = 0. \quad (8.8)$$

If we let U_A be the wave function corresponding to the bottom of the valence band, with energy E_A , and write

$$\Psi = U_A \phi, \quad (8.9)$$

Eq. (8.8) becomes

$$U_A \left[\frac{\hbar^2}{2m} \Delta \phi + (E - E_A) \phi \right] + 2 \nabla U_A \cdot \nabla \phi = 0. \quad (8.10)$$

The wave function, ϕ , corresponds to a free electron with normal mass and energy $E - E_A$ if $\nabla U_A \cdot \nabla \phi$ is negligible. This is roughly the case in ordinary metals, where $\nabla U_A = 0$ over most of the volume, except near the nuclei.

The extension of this approximation to alloys differs somewhat, depending on the type of solute atoms considered (Mott, 1936; Daniel, 1959; Roth, 1956; Blandin, 1960).

Interstitial solute atoms. To solve the equation

$$\frac{\hbar^2}{2m} \Delta \Psi + (E - V_A - V_p) \Psi = 0,$$

use is made of Eq. (8.9), hence

$$U_A \left[\frac{\hbar^2}{2m} \Delta \phi + (E - E_A - V_p) \phi \right] + 2 \nabla U_A \cdot \nabla \phi = 0. \quad (8.11)$$

If $\nabla U_A \cdot \nabla \phi$ is negligible, ϕ represents a free electron with normal mass and energy $E - E_A$, scattered by the perturbing potential V_p . The change from Eq. (8.10) is the introduction in ϕ of a scattered wave, with its main amplitude on the interstitial atom, where $\nabla U_A = 0$. The free-electron approximation should therefore be nearly as good for interstitial solutions as for the corresponding pure solvent, at least for small interstitial atoms (H, C, etc.).

Substitutional solute atoms. The neglect of $\nabla U_A \cdot \nabla \phi$ in Eq. (8.11) would be serious in this case, especially in treating the cohesive properties. Thus, because of the nuclear contribution to V_p , $\nabla \phi$ is large near the nucleus of the solute atom, where U_A oscillates strongly. It is then better to use Eq. (8.9) in the solvent and

$$\Psi = U_B \phi$$

in the atomic polyhedron of the solute atom B, where $U_B(E_B)$ is the wave function of the bottom of the valence band in a metal composed of pure B but having the same structure as A. If the perturbation V_p in the matrix is neglected, ϕ is then given by the solution to the equation

$$U_i \left[\frac{\hbar^2}{2m} \Delta \phi + (E - E_i - \phi) \right] + 2 \nabla U_i \cdot \nabla \phi = 0, \quad (8.12)$$

where $i = A$ in the matrix A and $i = B$ in the solute atoms B. When $\nabla U_i \cdot \nabla \phi$ is neglected, ϕ represents a free electron with normal mass and energy $E - E_A$ scattered by the solute atom having a square potential well of depth $E_B - E_A$. In such a potential, the scattered part of ϕ has a vanishing gradient at the center of the well and over most of the region where U_A oscillates.

This approximation should therefore be as good for the alloy as for the pure solvent when $E_A - E_B$ is not too large, so that the scattered part of ϕ does not vary strongly near the nucleus of the solute atom and V_p is negligible in the solvent. This condition is realized in practice if the constituents of the alloy do not differ greatly in valency. Whether or not the constituents come from different rows in the Periodic Table is not very important because

this does not greatly affect the energy of the bottom of the valence band. This approach has been justified in detail by Roth (1956).

Displacement of charges in the free electron model. For free electrons scattered by a central field of limited range, $V_p(r)$, the charge displaced can be computed exactly. The following results are of interest for cohesion:

(1) The total charge locally displaced by V_p is equal to

$$\frac{1}{\pi} \sum_{\sigma} \sum_l (2l + 1) \eta_l(E_M)$$

where $\eta_l(E_M)$ is the "phase shift" of the l^{th} spherical harmonics at the Fermi level E_M . Summation is on l and on the spin directions σ . This rule fixes a useful condition on the scattering potential to be employed. Furthermore, the change in density of states at the Fermi level produced by alloying is

$$\frac{N}{\pi} \sum_{\sigma} \sum_l (2l + 1) \frac{d\eta_l}{dE_M}$$

for small numbers N of solute atoms per unit volume.

(2) Most of this charge (usually more than 90 percent) resides within the atomic volume of the solute atom. This shows that in metallic alloys, ionic character is small (Daniel, 1959).

(3) Slow electrons are more strongly displaced than fast ones. This shows that, contrary to what we might think, the detailed electronic structure of a solute atom B is very different from what it would be in the pure metal B (Daniel, 1959).

(4) Outside the limited range of the perturbing field, where $V_p \cong 0$, interference effects produce the small displaced charges with alternating signs already described. The electron wave functions with wave number k can be analyzed in terms of spherical harmonics which must be linear combinations of spherical Bessel and Neumann functions:

$$j_l(kr) \cos \eta_l - \eta_l(kr) \sin \eta_l$$

where the phase shift, η_l , is a function both of V_p and r , and vanishes with V_p . The change in the charge density on alloying, for the electrons with wave number k , is then proportional to

$$[j_l(kr) \cos \eta_l - \eta_l(kr) \sin \eta_l]^2 - j_l^2(kr).$$

At large values of r , this expression decreases as r^{-2} times a sinusoidal function of kr .

These long-range oscillations do not contribute to the screening

of the solute atoms, because of their alternating signs. The contributions from all the wave numbers from 0 to the Fermi number k_F , compensate each other at large distances, so as to produce total oscillations which decrease at least as rapidly as r^{-3} ; in "ordinary" alloys, these oscillations are small even on the solvent atom nearest to a solute atom (Daniel, 1959).

Virtual bound states. In alloys of transition metals (e.g., Cu-Mn), the incompletely filled d shells give rise, by resonance with the conduction states, to resonance or virtual bound d levels. The thermal, electrical and magnetic properties of these alloys are well explained quantitatively by means of a simplified model of free conduction electrons scattered by an attractive potential; this gives rise to a virtual bound level, provided that the exchange terms within the virtual bound level are assumed to be the same as for the corresponding d shell, except for a quenching of the orbital moment. Resonance also broadens these states in space, thereby producing especially strong oscillations of charge of the kind described in the preceding paragraph.

8.2.4. Tight-Binding Approximations

In alloys with strong covalent character, the valence bond theory should be employed, perhaps using the Heitler-London localized atomic orbitals in alloys with strong ionic character, but more probably the Mulliken molecular orbitals. In fact, in a solid, resonance between the bonds of the various lattice cells broadens the bonding and antibonding orbitals into bonding and antibonding bands, which should be analyzed in terms of the tight-binding approximation. This broadening is usually larger than the splitting of the molecular states due to lattice symmetry; consequently, point group theoretical methods are not adequate when dealing with the various bonds made by an atom. Tight-binding equations for the entire alloy are usually necessary. As a result, theorems such as the Jahn-Teller effect on degeneracy are not always applicable.

8.2.5. Better One-Electron Approximations

It would be of great interest to develop a model which is capable of treating large perturbations, such as arise when the alloy constituents come from different periods, without requiring a free-electron assumption. Little progress has been made in this direction; results have been fragmentary and none are directly applicable to the cohesion problem:

- (1) It is a relatively simple matter to compute exactly the band

structure of one-dimensional models having a finite number of impurities and finite or infinite lengths. Randomly distributed impurities can also be treated by approximate methods.

(2) The effect of a local perturbation on the electron density of a one-dimensional lattice can also be treated exactly, both for full and incompletely full bands (Leman, 1960).

(3) For three-dimensional lattices, Roth (1956) has estimated numerically the effect on the electrical properties which would result if the effective mass of the matrix were different from the normal mass. Her method is, however, fairly cumbersome. The effects, both computed and observed, are relatively small and justify, on the whole, the use of "free electrons" for crude estimates in ordinary alloys.

(4) Finally, for any perturbed (nonperiodic) potential, the generalized expression for the change in the density of states produced by alloying is:

$$\frac{N}{\pi} \sum_{\sigma} \frac{d\eta}{dE_M},$$

and this change is produced by charges which are highly localized on the perturbing potential V_p . Here η is the phase shift associated with the S matrix for the scattering of Fermi electrons by V_p (Blandin, 1959; Roth, 1956). The general properties of $\eta(E)$ should, however, be known before such a relation can be really useful.

8.2.6. Size Effects

A classical elastic model is often used. Ways of treating the problem from a wave mechanical point of view will be discussed.

Classical elastic model. A solute atom B is represented by a sphere of elastic material inserted within a spherical cavity in the solvent material A. Both solute and solvent are treated as classical elastic media which have the same elastic constants as the pure macroscopic constituents; the normal radius of the sphere, r_B , is assumed to be the atomic radius of the pure solute element. The normal radius, r_A , of the spherical cavity is taken as the atomic radius of the solvent in substitutional alloys, but is assumed to be very small in interstitial alloys.

This model assumes essentially that:

(1) the electronic structures of the matrix and of the solute atom are only slightly or insignificantly perturbed by alloying. The preceding discussion shows that this should be a better approximation for the solvent than for the solute atom.

(2) the atomic arrangement is sufficiently close-packed for atomic polyhedra to be replaced by spheres. This will be a much better approximation for substitutional solutes in close-packed metals than for interstitial solutes or for covalently bonded solvent elements.

(3) the size effects are small enough to be treated by linear elasticity theory. This will often be acceptable for the solvent where the greater part of the energy is usually stored.

(4) the size effect is nevertheless large enough to be predominant over other factors. This is rarely true when applied to the cohesive energy, but might be a better assumption as far as average parameters and elastic constants are concerned.

Wave mechanical approach. Varley (1954) and Roth (1956) first pointed out that the size and valency factors do not act independently in influencing the properties of alloys. Blatt (1957) has shown that when a solute atom increases the volume of the cavity in the solvent by Δv , this repels locally an amount of positive (nuclear) charge:

$$\Delta Z = \frac{\Delta v}{v} p, \quad (8.13)$$

where $\frac{p}{v}$ is the number of valence electrons per unit volume in the solvent. In the classical model outlined above,

$$\frac{\Delta v}{v} = 6 \frac{r_B - r_A}{r_B + r_A} \frac{x}{x + (x_B/x_A)} \quad (8.14)$$

where x_A and x_B are the respective compressibilities of the two media; x is defined by:

$$x = \frac{1 + \nu}{2(1 - \nu)} \quad (8.15)$$

where ν is Poisson's ratio for the solvent. Subtracting ΔZ (Eq. (8.13)) from the difference of valency, Z , which must be screened, leads to a substantial improvement in the calculated electrical resistivities for "ordinary" alloys. Depending on the relative signs of Z and ΔZ , the size effect actually increases or decreases the electrical resistivity.

In the computation of the cohesive properties, the valency and size effects should similarly be treated together. Thus, the self-consistent potential which satisfies the total screening condition

$$Z + \Delta Z = \sum_{\sigma} \sum_l (2l + 1) \eta_l (E_M)$$

should be evaluated.

A variational method, minimizing the total energy for free electrons scattered by a perturbing square well potential, leads indeed to positive size factors for Cu-Ga, Cu-Ge and Cu-As alloys and negative ones for the Cu-Ni and Cu-Co systems (Thibaudier). The size factors obtained are, however, somewhat too large, and are also negative for Cu-Zn alloys, perhaps because of the roughness of the correlation terms used (Thomas-Fermi, see sec. 8.2.1). In such computations, only the strain energy of the solvent is computed by classical elasticity; strain and valence effects in the solute atoms are treated together by quantum mechanical methods.

8.3. Energies of Primary Solid Solutions

8.3.1. General Considerations

The solution (internal) energy at infinite dilution is the energy change when a B atom is taken from the pure element and introduced into the lattice of pure A. Little in general can be said about the energy of solution, except that it is usually small. It is fair to say that attempts to compute this energy may succeed in a few favorable cases, but fail for most alloys (Friedel, 1954, 1958; Varley, 1954; etc.). The only exceptions are the energies of solution, i.e., formation, of vacancies and interstitials (Huntington and Seitz, 1942; Fumi, 1955; Seeger, 1956; etc.). This is because these energies are large and the correlation terms are not important.

Solution energies are difficult to compute because they usually involve the sum of many small and compensating terms. Considering an ordinary substitutional alloy; for instance, let us first assume that A and B have the same lattice structure. If a B atom is introduced with its normal electronic structure, its electronic wave functions would fit fairly well with those of the surrounding matrix A, except that there will be some differences in the energies, E_A and E_B , corresponding to the bottom of the bands, and also in the Fermi energies, E_{MA} and E_{MB} . The first effect will lead to a scattering of the valence electrons by the square potential well of depth $E_B - E_A$ (see sec. 8.2.3). In ordinary (i.e., metallic) alloys, it is somewhat analogous to, but smaller than, the scattering by a free surface; in covalently-bonded structures, it corresponds to partial breaking of the covalent bonds. It gives rise to a positive surface energy term. The second effect leads to a transfer of electrons from the more electropositive constituent to the more electronegative; this introduces a negative stabilizing term. Both terms are proportional to the square of the respective energy differences, when these energy differences are not too large. Both

terms are therefore likely to be small, but have not been estimated accurately, mainly because the changes in the correlation energy associated with the solution of B in A are difficult to estimate.

If pure A and B have different lattice structures, the change in energy due to the allotropic transformation of pure B should be taken into account. This positive term is usually unknown and has rarely been estimated theoretically.

Finally, if all these contributions are negligible, when compared to the size factor, the elastic model described in sec. 8.2.6 gives a positive term of magnitude

$$\Delta E_s = \frac{6\pi r_A (r_B - r_A)^2}{\chi_B + \chi\chi_A}$$

or

$$\Delta E_I \cong \frac{3\pi r_B^3}{\chi_B + \chi\chi_A} \quad (8.16)$$

for substitutional and interstitial solutes respectively. Measurements on numerous alloys, however, have shown that the size factor is very rarely the dominant one.

8.3.2. Solution Entropy

The solution entropy at infinite dilution can be broken down into a positive term, corresponding to complete randomness, and a vibrational entropy term.

According to the experimental Kopp-Neumann law, the vibrational entropy of solution is small. The values measured are usually positive; they can then be explained by two contributions, each of the order of 10^{-4} ev per atomic percent of solute:

(a) *Strain energy*: Eqs. (8.16) are functions of the elastic constants and lattice parameters, which vary with temperatures. They therefore represent free energy terms. The corresponding entropy terms are:

$$\Delta S_{s,I} = \frac{\partial \Delta E_{s,I}}{\partial T} = y \Delta E_{s,I} \quad (8.17)$$

where

$$y = \frac{dx}{\chi dT} - \frac{3dr}{rdT} > 0. \quad (8.18)$$

(b) *Elastic constants*: Alloying changes the elastic constants and thus the phonon spectrum. Above the Debye temperature, θ , the corresponding entropy contribution is:

$$\Delta S_p = -3k \ln \frac{\theta}{\theta_B^c \theta_A^{1-c}}, \quad (8.19)$$

where θ_A and θ_B are the Debye temperatures of the pure elements and c is the solute concentration. Since $\theta \propto (r/\chi)^{1/2}$,

$$\Delta S_p = \frac{3k}{2} \left[\left\{ \frac{d\chi}{dc} - \frac{\chi_B - \chi_A}{\chi_A} \right\} - \left\{ \frac{dr}{r dc} - \frac{r_B - r_A}{r_A} \right\} \right]. \quad (8.20)$$

The first term in Eq. (8.20) is usually predominant and positive.

8.3.3. Interactions Between Solute Atoms

In metallic alloys, interactions between solute atoms are small. They can be either attractive, thus leading at low temperature to the formation of solute-rich phases, or repulsive, leading to local "short-range order" or even long-range order. Two general and fairly obvious theorems are often employed.

(a) If the interactions are only of short range, an attraction or a repulsion between solute atoms corresponds, respectively, to a negative or a positive curvature of the (internal) energy of formation of the alloy as a function of atomic concentration. This theorem should certainly apply to the purely electronic factors which can produce appreciable interactions only between near neighbors (sec. 8.2). It is indeed observed to apply in many alloy systems.

There are, however, experimental anomalies which are probably due to a peculiarity of the size effect, namely, that a solute atom, when dissolved well within the bulk of the material, produces in it a (infinitely small) uniform dilatation. This reduces the work necessary to introduce a second atom, and thus corresponds to an interaction energy, but without any interaction force if the solute and solvent have the same elastic constants. Under these conditions, the elastic term in the energy of formation of the alloy has a negative curvature given by

$$\Delta E_{el} = c(1 - c) \Delta E_{e,l}. \quad (8.21)$$

If $\chi_A \neq \chi_B$, the curvature of $\Delta E_{el}(c)$ is still strongly negative, while there is a small interaction force, varying as r^{-4} , which is either attractive or repulsive, depending on whether χ_B is greater than or less than χ_A (Eshelby, 1956). The whole elastic model, however, breaks down at distances short enough for this force to be noticeable.

(b) If the interactions are only between first neighbors and if they are additive (Ising model), a quantitative relation may exist between the energy of formation of the alloy and the energy of interaction between solute atoms, i.e.,

$$E = c(1 - c) pw, \quad (8.22)$$

where $2p$ is the number of neighbors and $w = 2(AB) - (AA) - (BB)$, in which (AB) denotes the energy of an AB bond, (AA) that of an AA bond, etc. This theorem is restricted to substitutional alloys consisting of two metals having the same structure.

Quasi-chemical approximations of this kind should apply only to covalently-bonded structures having little ionic character and in which size effects are small. Indeed, for most metallic alloys, Eq. (8.22) has no real validity.

Apart from these theorems, few general predictions can be made concerning the interactions between solute atoms in alloy systems. The interactions should be small, especially in transition metal solvents, and this seems to be borne out experimentally. They should also be of short range and therefore of appreciable magnitude only when they involve nearest neighbors. Attempts to predict the sign of these interactions theoretically (Friedel, 1954, 1958; Lazarus, 1954; March, 1956) have met with little success because, being based on the Thomas-Fermi approximation, they neglect the important long-range oscillations described above (sec. 8.2.3). In alloys of the transition elements having simple metallic structures (e.g., Cu-Fe), the strong magnetic coupling between solute atoms in neighboring lattice positions stabilizes the formation of solute-rich zones or clusters. This coupling originates through an overlap of the virtual bound d levels (Blandin and Friedel, 1959).

8.3.4. Stability Range of Primary Solid Solutions

This important technological property of alloy systems is difficult to analyze because it depends on the stability of the primary solid solution relative to that of any intermediate alloy phases or compounds which may exist in the system. Plots of electronegativity as a function of atomic radius have sometimes been used to predict which elements will dissolve appreciably in a given solvent (Darken and Gurry, 1952). A large difference in electronegativity leads to restricted primary solid solubilities (sec. 8.3.1) but stabilizes intermediate phases or compounds having a strongly ionic character. Electronegativity is, however, difficult to define precisely, so that this criterion is at best highly qualitative.

A large difference in atomic size also limits the extent of primary solid solubility. The primary solution becomes unstable and cannot exist at high concentrations if $\Delta E_{el}(c)$ is larger than kT . For an atomic concentration, c , less than 0.5 and $T < T_m$, T_m being the melting point of the alloy, Eqns. (8.20) and (8.16) predict, in

agreement with the Hume-Rothery rules, that extensive primary solid solubility will not occur if the atomic radii of the two components differ by more than about 15 percent.

8.3.5. Allotropic Transformations

The size factor seems to play an important role in at least two situations:

Interstitial alloy phases. There is no doubt, for example, that carbon stabilizes the austenite phase because it finds larger interstitial sites in face-centered cubic γ -iron than in the body-centered cubic α -phase.

Melting. If the strain energy is assumed to disappear on melting, and if that part of the free energy of melting which is not due to randomness varies linearly with concentration (Wagner, 1954), the liquidus, $c_L(T)$, and solidus, $c_s(T)$, are given, at small concentrations, by

$$\ln \frac{c_L}{c_s} = \frac{\Delta E - L_B(1 - T_A/T_B)}{kT_A}, \quad (8.23)$$

with

$$c_L - c_s = (1 - T/T_A) \frac{L_A}{kT_A}.$$

In these expressions, T_A and T_B are the melting temperatures, L_A and L_B the latent heats of melting of the pure components, and ΔE the strain energy associated with the presence of B in solid solution in A.

These relations are well obeyed in many alloy systems. They predict that the strain energy should increase with increasing departure of the liquidus and solidus curves from the straight line joining the melting temperatures of the two components in the binary equilibrium diagram. A fairly accurate idea of the size factor can therefore be obtained by mere inspection of the phase diagram.

8.4. Energies of Intermetallic Phases

The various factors listed in section 8.1 can play a role in determining the energies or structures of intermediate phases in alloy systems.

8.4.1. Ionic and Covalent Factors

The ionic and covalent factors lead to highly stoichiometric and ordered compounds. These factors are large only if semimetals, heavy metals, semiconductors or metalloids are involved as com-

ponents. In systems involving good metals, the ionic factor is small because the electronegativities of the component elements are comparable. Indeed, simplified models predict transfers of only fractions of electronic charges, from the elements with the higher to the lower valencies. These transfers are especially small in transition metal systems. Metals are also usually close packed and exhibit little tendency to form homopolar bonds. In alloys made of "good" metals, only the body-centered cubic phase observed in systems involving elements from the middle of each transition series (e.g., Cr, Mo, W) can be considered to have a strong covalent character.

8.4.2. Size Factor

The size factor also leads to highly stoichiometric and ordered alloys having approximately close-packed structures. Laves phases of the type A_2B are a typical example.

8.4.3. Electron Compounds

Purely metallic bonding in ordinary alloys leads to the stabilization of Hume-Rothery electron compounds, which occur at or near compositions having definite ratios of valence electrons per lattice site. As pointed out by Jones, these ratios are those for which the Fermi surface of a nearly free electron gas, moving in the average periodic potential of the alloy, would touch a well-defined Brillouin zone boundary. This corresponds to a particularly stable situation because of the depression of the electronic energies near the zone boundaries.

The success of this idea in explaining the existence of structurally analogous phases in many alloy systems leaves little doubt that it is at least roughly correct. Three characteristics of the Hume-Rothery phases may be noted:

(1) They are stable over a limited range of concentrations near certain critical electron/atom ratios. These intermetallic phases therefore exhibit a certain range of homogeneity, usually about 5 to 10 atomic percent.

(2) The stability does not require any special order; the phases are usually disordered, at least at high temperatures.

(3) The stability is not very pronounced. The energies of formation, as well as the liquidus and solidus, vary fairly smoothly with composition across the entire phase diagram. Indeed, some Hume-Rothery phases only occur because they are stabilized by other factors (e.g., order at low temperatures and vibrational entropy at high temperatures in β -brass).

Some restrictions must also be kept in mind. For example, Jones' explanation is only suitable for the conduction band, not for the d band of the transition metals (the electrons must be nearly free). Even then, the treatment given by Jones is only approximate, being based on the rigid-band approximation for alloys. Other factors (size, electronegativity) also complicate the picture.

8.5. Lattice Constants

8.5.1. Infinite Dilution

Thibaudier's *a priori* computations have been mentioned previously (sec. 8.2.6). The classical elastic model leads to a small deviation from the linear Vegard law, related to differences in compressibility between solute B and solvent A. More precisely, the relative variation of the lattice parameter with alloying is given by

$$\frac{r - r_A}{r_A} = \frac{r_B - r_A}{r_A} c \left[1 + \frac{x_A - x_B}{x_B + x x_A} \right] \quad (8.24)$$

where x is given by Eq. (8.15). The first term is Vegard's law. The second term introduces deviations, which are positive if the larger atoms are the harder ones, and negative otherwise. The initial slope dr/dc varies, according to this equation, from 0 for

$x_B \ll x_A$ to $(r_B - r_A)$ for $x_B = x_A$ and $(r_B - r_A) \left(1 + \frac{1}{x}\right)$ for $x_B \ll x_A$. The measured deviations are often in good semi-quantitative agreement with the model.

8.5.2. Anomalies for Zone Overlaps

In the rigid-band model, small anomalies in the slope of $r(c)$ are expected near critical electron/atom ratios, i.e., when the Fermi level overlaps from one Brillouin zone to the next. Anomalies of this type have been reported, for instance, in hexagonal Mg-base alloys. But the whole analysis is still highly controversial, and from a theoretical point of view, it depends critically on the form assumed for the Fermi surfaces and on the validity of the rigid-band model; on the experimental side, the observed effects, if any, are of the order of the accuracy of the measurements themselves.

8.6. Elastic Constants

8.6.1. Infinite Dilutions

The classical elastic model makes the following predictions concerning the variation of elastic constants with alloying:

(a) For solute and solvent with different atomic sizes but having the same elastic constants, the elastic constants of the alloy are independent of composition within the limits of linear elasticity. If higher-order terms are taken into account, a dilatation (due to interstitials or larger substitutionals) leads to a slight increase in the elastic constants on alloying, whereas a contraction (due to smaller substitutionals) reduces the elastic constants. This is the so-called Zener effect (Zener, 1949).

(b) For solute and solvent with the same sizes but different elastic constants

$$\frac{d(1/\chi)}{(1/\chi_A) dc} = \frac{1+x}{1+x(\chi_A/\chi_B)} \frac{(1/\chi_B) - (1/\chi_A)}{1/\chi_A} \quad (8.25)$$

and

$$\frac{d\mu}{\mu dc} = \frac{1+x}{(\mu_B/\mu_A) + x} \frac{\mu_B - \mu_A}{\mu_A}, \quad (8.26)$$

where x is again given by Eq. (8.15) and μ_A and μ_B are the shear moduli of the two pure components. By comparing the slopes at the two ends of the phase diagram, it is seen that both $(1/\chi)$ (c) and μ (c) exhibit a positive curvature, which is especially marked for the compressibility function. These variations are usually much larger than those given by the Zener effect. They are in good agreement with experiments on ordinary alloys (e.g., copper with Zn, Ga, Ge, or As as solute elements).

8.6.2. Anomalies for Zone Overlaps

Anomalies in the elastic constants at compositions corresponding to zone overlaps have been searched for in the same alloys used for lattice parameter studies, but have not been observed. This is not at all surprising, since the energy anomalies associated with zone overlaps are very small indeed.

8.7. Instructive Experiments

The essential conclusion of this survey is that the prospects for theoretical advances in this field do not appear very hopeful, at least until more systematic and accurate experimental data of the following types become available:

(a) Thermodynamic measurements on solid solutions, including measurements of the energy of solution at infinite dilution, of the variation in the energy of formation of the alloy as a function of concentration and measurements of the vibrational entropy term; for intermediate phases, data are also needed on the energies and entropies of formation.

(b) These measurements should preferably be supplemented by measurements of lattice parameters and elastic constants.

(c) More accurate determinations of liquidus and solidus curves should be made, especially in the dilute concentration ranges.

(d) Size effects and local order should be further studied by X-ray and neutron diffraction techniques, as well as by other means (specific heat, electrical resistivity, thermoelectric power, internal friction).

(e) In alloys with transition or heavy metals, a knowledge of the magnetic properties is essential for an understanding of their cohesion.

(f) A detailed knowledge of the electronic structure of the solute atoms and their surroundings would be very valuable. Optical, X-ray, or Knight shift measurements, both on the solute atoms and on the surrounding matrix, may give pertinent information in this regard.

8.8. References

- Blandin, A. and Friedel, J., 1959, *J. de Phys.*, **20**, 160.
 Blandin, A., 1960, Thesis, Ecole Nationale Supérieure des Mines, Paris.
 Blatt, F. J., 1957, *Phys. Rev.* **108**, 285.
 Bloembergen, N. and Rowland, T. J., 1953, *Acta Met.* **1**, 371.
 Cohen, M. H. and Heine, V., 1958, *Adv. Phys.* **7**, 395.
 Daniel, E., 1959, *J. Phys. Chem. Solids*, **10**, 174.
 Darken, L. S. and Gurry, R. W., 1953, *Physical Chemistry of Metals*, (McGraw-Hill, New York).
 Eshelby, J. D., 1956, "The Continuum Theory of Lattice Defects," in *Solid State Physics* (New York, Academic Press), Vol. 3, pp. 79-144.
 Friedel, J., 1952, *Phil. Mag.*, **43**, 153.
 ——— 1954, *Adv. Phys.*, **3**, 446.
 ——— 1956, *Can. J. Phys.*, **34**, 1190.
 ——— 1958, *Nuovo Cimento*, **2**, 287.
 Fumi, G., 1955, *Phil. Mag.*, **46**, 1007.
 Hume-Rothery, W., 1931, *The Metallic State* (Oxford, Clarendon Press).
 Huntington, H. B., 1942, *Phys. Rev.*, **61**, 325.
 ——— and Seitz, F., 1942, *Phys. Rev.* **61**, 315.
 Jones, H., 1934, *Proc. Roy. Soc. A* **144**, 255.
 Langer, J. S. and Vosko, S. H., 1960, *J. Phys. Chem. Solids* **12**, 196.
 Lazarus, D., 1954, *Phys. Rev.* **93**, 973.
 Leman, G., 1960, *J. Phys. Chem. Solids*, in the press.
 March, N. H., 1956, *Adv. Phys.* **6**, 1.
 Mott, N. F., 1935, *Proc. Phys. Soc.* **47**, 571.
 ——— 1936, *Proc. Cam. Phil. Soc.* **32**, 281.
 ——— and Jones, H., 1936, *Metals and Alloys* (Oxford, Clarendon Press).
 Muto, T., 1936, *Scientific Papers Inst. Physical and Chemical Res. Tokyo* **30**, 99.
 Nordheim, L., 1931, *Ann. Physik* **9**, 641.

- Roth, L., 1956, Thesis, Radcliffe College, Cambridge, Massachusetts; Technical Report No. 267, Cruft Laboratory, Harvard University.
- Seeger, A. and Bross, H., 1956, *Z. Phys.* **145**, 161.
- Thibaudier, G., to be published.
- Varley, J. H. O., 1954, *Phil. Mag.* **45**, 887.
- Wagner, C., 1954, *Acta Met.* **2**, 242.
- Zener, C., 1949, *Acta Cryst.* **2**, 163.

Part III

MAGNETISM AND MAGNETIC MATERIALS

Chairman

Dr. J. E. GOLDMAN
*Ford Motor Co.,
Dearborn, Mich.*

Members

Dr. P. W. ANDERSON
*Bell Telephone Laboratories
Murray Hill, N. J.*

Dr. A. M. CLOGSTON
*Bell Telephone Laboratories
Murray Hill, N. J.*

Dr. CHARLES P. BEAN
*General Electric Co.
Schenectady, N. Y.*

Dr. WALTER MARSHALL
*Atomic Energy Research
Establishment
Harwell
Didcot, Berks. England*

Prof. W. F. BROWN, Jr.
*University of Minnesota
Minneapolis, Minn.*

Dr. R. J. WEISS
*Watertown Arsenal
Watertown, Mass.*

CONTENTS

	Page
1. INTRODUCTION	76
2. PARAMAGNETISM AND PARAMAGNETIC MATERIALS.....	79
2.1. Paramagnetic Studies	80
2.2. Paramagnetic Materials Technology.....	83
3. COOPERATIVE MAGNETIC INTERACTIONS.....	86
3.1. The Fundamental Origins of Exchange Interactions.....	86
3.1.1. Salts	86
3.1.2. Dilute Alloys	94
3.1.3. Rare Earth Metals.....	96
3.1.4. Transition Metals	97
3.2. Experimental Observation of Exchange Effects.....	100
3.2.1. Curie or Néel Temperature.....	100
3.2.2. Total Energy of Magnetization.....	101
3.2.3. Low Temperature Magnetization and Low Temperature Specific Heat	101

3.2.4. Susceptibility at the Néel Temperature	108
3.2.5. Domain Wall Thickness	108
3.2.6. Inelastic Neutron Scattering	104
3.2.7. Paramagnetic, Spin Wave, and Ferromagnetic Resonance	104
3.2.8. Resistivity of Magnetic Materials	105
3.2.9. Perpendicular Susceptibility Below the Néel Temperature	106
3.2.10. Antiferromagnetic Resonance	107
4. STATIC PROPERTIES OF FERROMAGNETIC AND FERRIMAGNETIC MATERIALS	107
4.1. Intrinsic Properties	107
4.1.1. Magnetization	107
4.1.2. Experimental Determination of Spontaneous Magnetization	111
4.1.3. Magnetocrystalline and Magnetoelastic Effects	116
4.2. Structure-Sensitive Properties	119
4.2.1. Domains and Domain Walls	120
4.2.2. The Magnetization Process	123
4.2.3. Micromagnetics	125
4.2.4. Experimental Techniques	128
5. DYNAMIC PROPERTIES OF FERROMAGNETIC AND FERRIMAGNETIC MATERIALS	129
5.1. Ferromagnetic Resonance	129
5.1.1. Ferromagnetic Insulators	130
5.1.2. Ferromagnetic Metals	135
5.2. A. C. Losses	138
5.3 Magnetic After Effects	139
6. DIAMAGNETISM AND SUPERCONDUCTIVITY	142
7. BIBLIOGRAPHY AND REFERENCES	147
7.1. Bibliography: Books and Review Articles	147
7.2. References	148

MAGNETISM AND MAGNETIC MATERIALS

1. INTRODUCTION

Few areas in the history of modern materials science demonstrate as vividly the interplay between basic science and materials technology as does the field of magnetic materials. Many examples in magnetic materials research and technology make amply evident the two methods by which important advances occur: the evolutionary improvement in the properties of technical materials made possible by the continuously increasing state of theoretical knowledge concerning the origin of these properties; and the revolutionary advances, founded upon theoretical or experimental discoveries, on the basis of which new classes of materials are in-

vented or new uses are found for old materials with improved properties.

Before World War II, the exclusive domain of the science and technology of magnetic materials was, in effect, to extend the limits of coercive force (and permeability) at both ends of the spectrum, i.e., for both "soft" and "hard" magnetic materials. In the perspective of the state of the art—or science—as it then existed, improvements of many orders of magnitude were achieved in the magnetic properties of existing metals and alloys and new alloy combinations were developed, primarily as a result of work conducted in a handful of industrial research laboratories.

During the decade or so immediately preceding World War II, theoretical advances to parallel the important experimental ones suddenly began to flower. In relatively short order, people began to understand—at least qualitatively—the origin of exchange interactions, the origin of domains and some of their dynamics. This stage unquestionably provided the impetus for the major "revolution" which has taken place in the whole concept of magnetic phenomena and materials since the War. In rapid order came: the discovery of antiferromagnetism and ferrimagnetism, in the wake of which a completely new class of materials (oxides in various crystalline forms) was developed, which in turn pushed forward the utilizable frequency range for magnetic materials; the theory of single domain effects in fine particles and thin films and the theoretical prediction of unusual coercive properties of such materials; the discovery of neutron diffraction and magnetic resonance as tools for investigating atomic and nuclear magnetism; the evolution of the maser as a new and probably first application of paramagnetism in electronic technology, and the cryotron as a first application of diamagnetism (superconductivity).

In summary, where 20 years ago we were limited essentially to only one class of utilizable magnetic phenomena (ferromagnetism), today we are aware of the existence of five classes of useful magnetic materials (ferro-, antiferro-, ferri-, para-, and diamagnetic). At that time, interest in magnetic materials was focussed exclusively on metals and alloys; magnetic materials technology today extends to oxides such as spinels, garnets, rubies, and sapphires, to submicron particles and even to cadmium whiskers. Before World War II the number of active researchers in all aspects of magnetic materials research and development in the entire world could be estimated at approximately 100; today there must be of the order of 4,000. Two hundred papers in the field of magnetism and magnetic materials and devices were submitted to the 1960

Conference on Magnetism and Magnetic Materials, one of a series held annually in the United States.

The intent of this report is to examine the status of what we know about magnetism today and the implications of this present knowledge in terms of materials development potential of the near and longer-range future. Although the emphasis is placed on what is now known, some attention is given to the voids in the current state of knowledge.

It is hoped that what follows will help point up these voids and will suggest how filling them would affect the science and technology of magnetism and magnetic materials. Any subdivision of the subject matter is necessarily arbitrary and subject to criticism. In this presentation, we have devoted one section to paramagnetism and one to diamagnetism including superconductivity—recognizing the traditional classification. Three sections are devoted to cooperative magnetic phenomena, one dealing with fundamental interactions, one with static properties and one with dynamic properties. Certain conclusions are inescapable even before delving into detailed analysis: that there are certain glaring voids in the state of our knowledge of some very basic magnetic effects; that much remains to be learned concerning the nature of the wave functions of magnetic electrons, the exchange interactions and the ensuing second-order effects such as anisotropy.

Any advances made in the resolution of these phenomena in the near future must inevitably open up new vistas in our understanding and ultimately in our exploitation of magnetic phenomena. It should become equally clear that another avenue that must inevitably lead to new scientific and technological advances lies in the interaction of magnetism with other classical physical disciplines as in magneto-acoustic and magneto-optical phenomena. These are fields which are just beginning to grow and it is here that we may look for important new developments. There are some instances where theory has now just begun to overtake experimental information. The line width in garnets is one of these; superconductivity is another. The surface of potential exploitation of these phenomena has hardly been scratched and it remains to be seen what new materials or devices may arise or benefit from this newly acquired knowledge. But for the present, the crystal ball must be put aside while the subject is studied—perhaps to be picked up again to see where the future will lead us.

Before proceeding with the detailed discussions which follow, the authors wish to acknowledge the very helpful suggestions, comments and criticisms of Dean Harvey Brooks.

2. PARAMAGNETISM AND PARAMAGNETIC MATERIALS

Many materials of various types are paramagnetic. A list of even those materials exhibiting paramagnetic resonance would include such widely diverse substances as: biological materials (Commoner, Townsend, and Pake, 1954; Commoner, Heise, and Townsend, 1956); ionic crystals containing F-centers (Kip, Kittel, Levy, and Portis, 1953); free electrons in metals (Griswold, Kip, and Kittel, 1952); and semiconductors (Portis, Kip, Kittel, and Brattain, 1953); impurity centers in semiconductors (Fletcher, Yager, Pearson, Holden, and Merritt, 1954), and other valence crystals (Baker, Winslow, and Yager, 1955; Smith, Gelles, and Sorokin, 1959; Weeks, 1956); solutions (Hutchinson and Pastor, 1951; Tinkham, Weinstein, and Kip, 1951); and, finally, the materials more normally thought of as paramagnetic, namely, ions of the iron group, rare earth and actinide series, and organic free radicals. Paramagnetic resonance and spin Hamiltonians have been discussed in several important review articles (Bleaney and Stevens, 1953R; Bowers and Owen, 1955R).

In most instances, however, paramagnetism and particularly electron-spin resonance are used as tools to investigate other primary properties: e.g., in the study of impurity wave-functions in semiconductors (Feher, 1956), of F-center (Feher, 1957) and V-center (Kastner and Kanzig, 1957) wave functions in ionic solids, of spin susceptibility (Schumacher and Slichter, 1956) or electron motion (Feher and Kip, 1955; Dyson, 1955) in metals. The present study on the other hand need consider in detail only two categories: (a) paramagnetic studies throwing light on important properties of cooperatively magnetic materials, typically magnetically dilute materials which, when more concentrated, would be ferro- or antiferromagnetic; (b) paramagnetic materials in relation to technical uses as paramagnetics per se. A list of most of the possible uses, some of which are unlikely, is: (1) most important, 3-level masers (Bloembergen, 1956; Basov and Prokhorov, 1955); (2) the less practical two-level masers (Conbrisson, Honig and Townes, 1956; Feher, Gordon, Buehler, Gere and Thurmond, 1958) and parametric amplifiers (Anderson, 1957; Javan, 1957); also possible ultrasonic uses; (3) spin-echo storage (Gordon and Mowers, 1958); (4) magnetometry: the g-marker use of free radicals is well known. Paramagnetic resonance high-sensitivity magnetometry has been proposed; (5) paramagnetic resonance is nonreciprocal, and it was proposed for gyrator use along with ferromagnetic resonance; (6) adiabatic demagnetization.

Some of these last uses might be disposed of in a word. Where-

ever sharp lines and no demagnetizing fields are valuable one might as well go to the extreme of nuclear resonance (as in magnetometry). Similarly, where strength of resonance as well as sharp lines are useful—parametric amplifiers, gyrators, other magnetometry—optimization leads to increasing the exchange integral and concentration, and we find ourselves with a ferromagnet—especially now, with the availability of tenth-oersted lines in Y.I.G. samples. Thus, practical applications of paramagnetism will tend to be highly specialized.

2.1. Paramagnetic Studies

A major recent development has been the realization that much of ligand field theory (Orgel and Griffith, 1957R; Ballhausen, 1954R) applies as well to concentrated salts (ferrites, garnets, lanthanates, fluorides) as to isolated paramagnetic ions. Thus, the study of the isolated ions contributes to our knowledge of technical materials.

Ligand field theory is the theory of the isolated magnetic ion in relation to its environment of anions or "ligands" (typically H_2O , O^{2-} , F^- , etc.). The old point of view of the Hund-Van Vleck crystal field theory was that these ligands created an electrostatic field at the paramagnetic ion which split up the energy levels of the d -electron. Of the d -electron wave functions, those which point toward the ligands (like the ones with symmetry $x^2 - y^2$ or $2z^2 - x^2 - y^2$ (e_g), when the ligands form an octahedron) will be repelled by the negative charges on the ligands and raised in energy; those pointing between them (like xy , yz , zx in the octahedron (t_{2g})) will not be raised. The resulting splitting is called 10 Dq. Further, in some cases a distortion of the often approximately cubic surroundings can split these functions still further. In particular, a tetragonal stretching of the octahedron will split the two e_g levels. Table 1 shows, for the more usual octahedral surroundings (tetrahedral ligands reverse the role of e and t), the relevant facts about crystal field effects in iron-group ions. The examples of Table 1 may help to show what this means.

For the simplest case, examine the data for Cu^{2+} . There is one e_g hole. However, the so-called Jahn-Teller effect, of great importance in these substances, forces the lattice to distort when it can so that it removes this degeneracy. Thus, Cu^{2+} distorts tetragonally, leaving no orbital degeneracy, tetragonal symmetry and spin 1/2; spin-orbit effects give anisotropic g .

More complex is, for example, V^{3+} . Here the $e_g - t_{2g}$ theory predicts a 3-fold degeneracy for the one hole in the t_{2g} shell. The

TABLE 1
CRYSTAL FIELD EFFECTS IN IRON GROUP IONS

Field Effects	Ti ⁺⁺⁺	V ⁺⁺⁺⁺	Cr ⁺⁺⁺	Mn ⁺⁺⁺	Mn ⁺⁺ Fe ⁺⁺	Fe ⁺⁺	Co ⁺⁺	Ni ⁺⁺	Cu ⁺⁺
10 Dq Configuration in terms of cubic orbitals.	t_{2g}	$(t_{2g})^2$	$(t_{2g})^3$	$(t_{2g})^2e_g$	d^5	$(t_{2g})^4e_g$	$(t_{2g})^5e_g$	$(t_{2g})^6e_g$	$(t_{2g})^7e_g$
Orbital degeneracy	3	3	1	2	1	3	3	1	2
Spin	1/2	1	3/2	2	5/2	2	3/2	1	1/2
Orbit and Spin degeneracy	6	9	4	10	6	15			
Distortion from cubic	?	?	no	tetr.	no	?	?	no	tetra.
Usual Degeneracy			4	5	6			3	2

spin degeneracy gives a total of 6 states. In this, as in Ti^{3+} , Fe^{2+} and Co^{2+} , the orbital degeneracy leads to complex magnetic behavior which there is not time here to discuss. The general rule throughout the group, however, is that the orbital degeneracy is rather widely split by crystal field effects, and the ground state has relatively little orbital angular momentum (is "quenched"). Many important effects, such as g-shifts, relaxation, anisotropy, etc., depend very sensitively on the degree of "unquenching" which may occur because the spin is coupled to the orbits by the relativistic spin-orbit effect.

Modern points of view take the purely electrostatic nature of the crystal fields less seriously, emphasizing the fact that the crystal field theory is a convenient way of describing the symmetry of the situation, but that Dq really has components of several kinds. The ligand field ideas of Stevens and Orgel suggests that one component is contributed by the antibonding character associated with the magnetic orbitals, because the ligand wave functions are partially bonding ones. There are also effects of the modification of the d orbitals by screening the ligand functions, of orthogonalization (Phillips, 1959), and of attraction to the ligand cores. Finally, a group-theoretical analysis more complete than crystal field theory by Koster and Slater (1959) shows that the theory has, in most cases, fewer parameters than may be necessary to fit the data; "spin Hamiltonians" may have to be further generalized. In general, theoretical and experimental work in this area is very vigorous and will continue to be of great importance. A vital field is the direct study of wave function amplitudes by hyperfine interactions with the surrounding nuclei; recent data show that perhaps at least in fluorides the covalent effects are not large.

The ligand field theory has had remarkable qualitative success in explaining a variety of effects in paramagnets. Recently this work has been extended to ferromagnetic materials with considerable success. Obviously, the ion will have the paramagnetic anisotropy in the concentrated material which it had as an isolated ion. Lattice distortions due to the Jahn-Teller effect which occur in paramagnets can also be expected in ferromagnets (Dunitz and Orgel, 1957). Equally, hyperfine interactions, indicating the extension of the wave function to the ligands, appear to be the same in para- and cooperatively-magnetic materials (Tinkham, 1956; Shulman and Jaccarino, 1957), as do the optical spectra indicating the positions of the various ionic levels (Newman and Chrenko, 1959; Pratt and Coelho, 1959; Morin, 1958).

Cation distributions in spinels, etc., are influenced by the energy

contributions caused by the ligand field, as shown by Dunitz and Orgel (1957). Finally, the information which the ligand field theory of the isolated ion gives us as to the overlap of the magnetic electrons on nearby ions can be included in our ideas of the exchange effect itself (Goodenough, 1958; Wollan, Child, Koehler and Wilkinson, 1958; Anderson, to be published).

In spite of these successes, the ligand field concept itself, and especially the question of covalent effects, needs a great deal of clarification from both experiment and theory. The increasing ability of neutron diffraction to give details of magnetic wave functions will provide vital information, as will more detailed and accurate hyperfine structure studies, and the expanding field of optical spectroscopy.

It might be mentioned that a field which may be equally interesting, but which has received less attention, is the study of the paramagnetism of diluted metal systems (Owen, Browne, Knight and Kittel, 1956; Owen, Browne, Arp and Kip, 1957; Schmitt and Jacobs, 1957; Matthias, Suhl and Corenzwit, 1958). Where the solute has a magnetic moment, the strength of exchange interaction can be determined from the temperature dependence of susceptibility. From the concentration dependence, it is possible to determine the range of such interactions. In particular, the onset of a cooperative phenomenon with increasing concentration of solute is directly related to the range of interaction and the coordination of solute atoms (Sato, Arrott and Kikuchi, 1959).

In a second type of dilution study, paramagnetic resonance has been used to study pairs of magnetic ions and to deduce both isotropic and anisotropic exchange integrals. Such studies could be extremely valuable for metals as well as for insulators and will be discussed later in connection with measurement of exchange integrals.

2.2. Paramagnetic Materials Technology

A whole new branch of materials technology has grown up based on the requirements of the solid state maser. Previously the paramagnetism of diluted ions was studied for the resulting light it shed on concentrated materials, or for its contributions to an understanding in some other field of physics. Now dilute systems are created for their paramagnetic behavior, and such diluents as ruby (Al_2O_3), $\text{K}_2\text{Co}(\text{CN})_6$, beryl, the metatungstates, MgO (F-centers), etc., have been treated extensively. The material requirements include: perfection of crystal structure (to avoid strain broadening); optimum relaxation times; absence of hyperfine in-

teraction with the diluent (again for narrow lines); doping possibilities; mechanical and chemical stability; and mechanical properties, such as machinability. On the other hand, it would be a mistake to confine interest, at least for research, to the optimum materials only, because the absence of hyperfine interaction limits our information about wave functions. Furthermore, the relaxation question is, as we shall see, far from resolved, and can probably be studied to greater advantage in substances with a wide range of relaxation times.

The output of the three-level maser is proportional to the relaxation rate (as we see by noting that for each quantum of pump power, one quantum each must be emitted and must relax). While a maser can, in principle, work when all relaxation rates are equal, its operation would be greatly facilitated by the manipulation of relaxation times, which so far has been achieved only by "doping" with a paramagnetic ion that has a line coincident with one of the working ones, and relying on the "cross-relaxation" process (Feher and Scovil, 1957). These facts have brought into technical prominence old and vexing questions about mechanisms of spin-lattice relaxation at low temperatures which, from a materials physics point of view, represent the most challenging area of the study of paramagnetism at present.

The first suggestion about relaxation came from Townes et al. (Gordmaine, Alsop, Nash and Townes, 1958) who pointed out that relaxation in these substances would be greatly impeded because the phonons which are emitted when a spin relaxes have difficulty in leaving the sample, both because the number of modes is too few and because they are rapidly reabsorbed by other spins. Their detailed theory, which included the idea that the phonon modes are greatly broadened because of the reabsorption process, was disproved by Kittel (1958) and Anderson (1959), and experimentally also does not work well. A more sophisticated theory based on the same ideas, together with the inclusion of spin-spin interactions to some degree, has been developed, but it seems at present that this "phonon trapping" mechanism does not have great importance in the usual materials of technical interest. Study of materials where phonon effects may be larger is an important area for future research.

Theories of cross-relaxation based entirely on magnetic spin-spin interactions have been also put forward (Bloembergen, (1959A)). It can be argued that this process will usually be faster than phonon cross-relaxations. Cross-relaxations at harmonic frequencies, probably caused in this way, also have been observed.

This area represents an opportunity for fruitful studies from all points of view: devices, materials, and theory. In particular, no complete theory of this type of cross-relaxation process exists. Probably the single most puzzling aspect of the problem, an aspect which appears in almost all relaxation measurements, is the rather strong increase of relaxation rate with concentration.

Obviously each theory applies in the proper domain; to us the clearest present need is for a general realization of the complexity of these processes, and the recognition that each set of ideas, no doubt combinations thereof and, still more, entirely new concepts—will be fruitful, too. There is no reason to suppose that the fundamental physics of the maser is any simpler than that of the gas tube or the transistor. It should also be kept in mind that this field will inevitably interact, if only from a measurement standpoint, with that of very high-frequency ultrasonics (Dransfeld and Bömmel, 1958, 1959) and now, with the discovery of the laser, with the field of optical absorption as well.

A last mention should be made of experiments on two-level maser action (Conbrisson, Honig and Townes, 1956; Feher, Gordon, Buehler, Gere and Thurmond, 1958; Chester, Wagner and Castle, 1958) which have left unanswered many interesting questions, some of which are relevant to the relaxation dilemma.

Adiabatic demagnetization is, as yet, the only technique known which allows the investigation of temperatures below 0.3°K . At one extreme, there is now commercially available a two-stage refrigerator using this technique, together with superconducting heat switches, which makes the region below 0.1°K available with relative ease. On the other hand, there is the recent achievement by Kurti and Simon (Kurti, Robinson, Simon and Spohr, 1958) of temperatures of 2×10^{-5} degrees K, by adiabatic demagnetization, and the well-known use of a refined version of this technique in the first experimental disproof of parity conservation (Wu, Ambler, Hayward, Hoppes, and Hudson, 1957). Still lower temperatures have now been achieved in Kurti's laboratory.

For refined experiments, such as the latter two, the choice of a working substance is dictated almost entirely by the particular experimental conditions required, as are the major technical difficulties. In general, a paramagnetic salt for extreme temperatures is chosen to be as magnetically dense as is compatible with the lowest temperatures desired, bearing in mind that the spin interactions (exchange or μ^2/r^3) must be smaller than kT . Small spin-values (Cu, Ti, Ce) are preferable because the loss in entropy $\ln(2S + 1)$ is more than compensated for by lower interactions

(going as $S(S + 1)$). Demagnetization by rotation, using anisotropic g 's as in Ce salts, is an interesting new technique aimed at starting demagnetization from a low temperature.

A related, growing use of paramagnetism is the alignment of nuclei by resonance techniques based on interchanging electronic for nuclear polarization (Overhauser Effect). Recent work has shown the feasibility of getting rather large proton polarization in special materials by a technique involving the saturation of certain "forbidden" cross-transitions and spin diffusion in the nuclear system. Materials studies in this field are beginning (Abraham and Proctor, 1958; Abraham, McCausland and Robinson, 1959). Another new method uses "hot" electrons (Feher, 1959).

3. COOPERATIVE MAGNETIC INTERACTIONS

3.1. The Fundamental Origins of Exchange Interactions

It is convenient to discuss, separately, exchange effects in the following materials; (1) salts, (2) dilute alloys, (3) rare earth metals, (4) transition metals. It is fair to say that our understanding of the origins of these exchange effects is quite unsatisfactory and therefore all parts of this summary are open to some doubt.

3.1.1. Salts

In salts it is generally agreed that the exchange effects may be represented by a Dirac vector model which gives the dependence of the energy upon the spin orientations as

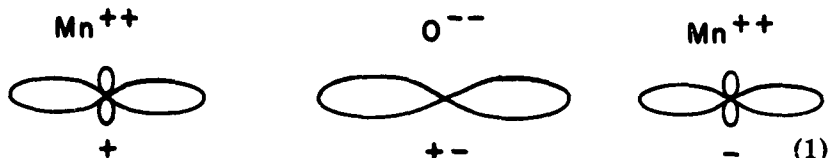
$$H = - \sum 2J_{ij} S_i \cdot S_j \quad (3.1)$$

where the sum runs over all pairs of spins S_i , S_j and the J_{ij} are exchange parameters whose origin we shall discuss. Clearly, positive J_{ij} corresponds to ferromagnetic coupling, and negative J_{ij} to antiferromagnetic coupling. The J_{ij} certainly depend on the overlaps between wave functions centered on different atoms and because these overlaps fall off exponentially with distance we may assume the exchange interactions are of very short range. Hence, it is always a good approximation to restrict the sum over spin pairs to just those pairs which are close together, and usually only nearest neighbor and next-nearest neighbor pairs need be included. In salts the magnetic ions are usually separated by nonmagnetic ions with closed shells, O^{2-} , F^{2-} , etc., and it seems likely that direct exchange is unimportant because the direct overlap between the wave functions of the magnetic electrons is so small. In such situ-

ations the spin coupling must arise from what is called superexchange.

A superexchange process is intimately connected with the (nominally) nonmagnetic anion and always involves a departure from the purely ionic (Heitler-London) state to an excited state in which the closed shell structure of the anion is destroyed. Superexchange has been discussed many times (Kramers, 1934; Anderson, 1950; Van Vleck, 1950, 1951; Pratt, 1955; Nesbet, 1958) but all these discussions are incomplete and to a considerable extent unsatisfactory. Nevertheless, it seems probable that we now understand the basic mechanisms which give rise to superexchange although we do not yet know the conditions under which one mechanism is more important than the other. Quantitative calculations of superexchange are very difficult, and nothing approaching a satisfactory calculation has yet been performed. We are therefore obliged to give a purely qualitative discussion.

For illustrative purposes let us first consider an explicit example, say O^{--} , as the anion with two cations, say Mn^{++} , on either side. We suppose for simplicity that only three orbitals are important, a $2p$ from the O^{--} and a $3d$ from each of the Mn^{++} . We shall indicate the electron occupation of these orbitals in the purely ionic state by means of the diagram.



if the Mn^{++} spins are antiparallel, and by



if the Mn^{++} spins are parallel. In these diagrams the drawing of the wave functions is, of course, only schematic and the + and - means that the $2p$ orbital of the O^{--} is doubly occupied, i.e., has two electrons, and the $3d$ orbitals are each singly occupied with either + or -. Of course each atom has other orbitals which are occupied, but, because these other orbitals point in directions away from the line joining the $Mn^{++} - O^{--} - Mn^{++}$ ions, they give only very small overlap effects which we may ignore. However, it is important to remember that the magnetic atoms have electrons in these other orbitals because the electrons in the particular orbitals we write down interact with those in the other orbitals of the same atom by "intra-atomic" exchange, i.e., according to Hund's rule.

We therefore look for effects produced by mixing small amounts of excited states into the ground states (1) and (2). The excited states that matter are those in which one or both of the electrons in the $2p$ orbital of O^{--} are transferred to the $3d$ orbitals of the Mn^{++} ions. If we transfer just one electron we can form two states from (1) which are

$$+- \qquad \qquad \qquad + \qquad \qquad \qquad - \qquad (3a)$$

and

$$+ \qquad \qquad \qquad - \qquad \qquad \qquad +- \qquad (3b)$$

and two states from (2) which are

$$+- \qquad \qquad \qquad + \qquad \qquad \qquad + \qquad (4a)$$

and

$$+ \qquad \qquad \qquad + \qquad \qquad \qquad -+ \qquad (4b)$$

If we transfer both electrons we can form one state from (1) which is

$$+- \qquad \qquad \qquad \qquad \qquad \qquad \qquad +- \qquad (5)$$

but we cannot form an analogous state from (2) because of the Pauli principle. We are reasonably confident that these excited states (3), (4) and (5) are the only ones which are important for superexchange but we do not yet know whether the "single transfer" process leading to (3) and (4) or the "double transfer" process leading to (5) is the dominant mechanism.

The single transfer process leading to (3) and (4) was first mentioned by Kramers (1934) and discussed in some detail by Anderson (1950). The essence of Anderson's argument is that the excited states (3) and (4) have slightly different energies because of the exchange coupling between the single electron left in the $2p$ orbital of O^{--} and the electron left unpaired on one of the Mn^{++} ions. If this coupling is of the expected antiferromagnetic sign, then the states (3) have lower energy than the states (4) and this leads to (1) having lower energy than (2). Hence, effectively, there is an antiferromagnetic coupling between the Mn spins. Anderson showed that the energy difference produced in this way between (1) and (2) was of order of magnitude d^2A/u^2 ; i.e.,

$$J \approx - \frac{1}{4s^2} \frac{d^2A}{u^2} \qquad (3.2)$$

where d is the one-electron transfer integral connecting (1) to (3), A is the exchange interaction between a $2p$ electron of the O^{--} and a $3d$ electron of Mn^{++} , and u is the energy of (2) relative to (1). It

seems plausible that all other 3rd-order contributions are small, although we cannot be sure of this; we shall not discuss them here.

The double transfer process leading to (5) appears to have been first suggested by Nesbet (1958); it had not been recognized in the earlier papers on superexchange. It gives an effective antiferromagnetic coupling in second-order perturbation theory which is of order of magnitude:

$$J \sim - \frac{1}{4s^2} \frac{b^2}{E} \quad (3.8)$$

where b is the "double transfer" matrix element connecting (1) to (5) and E is the energy of (5) relative to (1). To see the reason that this process gives antiferromagnetic coupling, recall that any admixture of an excited state lowers the energy of the ground state according to second-order perturbation theory. Hence, the double transfer process must lower the energy of (1) relative to (2); i.e., an effective antiferromagnetic coupling is produced.

This double transfer process gives a spin coupling in second-order perturbation theory whereas the single transfer process gives a spin coupling only in third-order perturbation theory. This does not necessarily mean that the double transfer process is dominant over the single transfer process because position in a formal perturbation theory is not a reliable guide to the order of magnitude of the effects. Indeed, we notice that in both cases the coupling is of fourth-order in the $2p$ - $3d$ overlap. Rough estimates by Keffer and Oguchi (1959) suggest that the single transfer process is several times larger than the double transfer process.

From this qualitative discussion several conclusions can be drawn. We notice that both transfer processes couple spins belonging to $3d$ orbitals which overlap strongly with a single p orbital. Now the strongest overlaps are given for the three ions arranged in a straight line as we have shown in these single diagrams; in particular, if the Mn^{++} ions were placed to make a 90° angle with the O^{--} ion we could not get large simultaneous overlaps with a single p orbital. Hence, we expect superexchange to couple strongly the spins of those magnetic ions which, seen from the anion, are at 180° from one another and to couple only weakly those ions which are at 90° . What experimental evidence is available seems to support this qualitative conclusion; e.g., the spin arrangements observed in MnO , FeO , NiO and CoO by neutron diffraction can only be understood in terms of a strong 180° coupling through the O^{--} ion.

Given this 180° rule it is also possible to postulate plausible rules

for the sign of the superexchange coupling. In the particular example $\text{Mn}^{2+} - \text{O}^{2-} - \text{Mn}^{2+}$ we have just considered, we concluded that the coupling would be antiferromagnetic, but that for other examples, different orbitals would be occupied, and we could expect a different result. A simple consideration of the single transfer and double transfer processes, remembering Hund's rule, then leads to the rules for superexchange:

- (a) If either orbital is doubly occupied, the coupling is negligible;
- (b) If each orbital has one electron, the coupling is antiferromagnetic;
- (c) If one orbital has one electron and the other is empty the coupling is ferromagnetic;
- (d) If both orbitals are empty, the coupling is antiferromagnetic.

Rule (a) is obvious and rule (b) is illustrated by the example $\text{Mn}^{2+} - \text{O}^{2-} - \text{Mn}^{2+}$ we have discussed in detail. It is interesting to note that rule (b) is given by both the single transfer and double transfer process. Rule (c) is also given by both processes but rule (d) is given only by the double transfer process because the single transfer process predicts negligible coupling in this case. It follows that we may expect the coupling in case (d) to be much weaker than for (b) and (c). It is suggestive that both processes lead to the same rules; this has led Anderson to formulate a theory in which both appear as part of the same process.

But these four rules are incomplete because we have not yet discussed what is called "double exchange" and a related process we shall call "virtual double exchange." These processes are important when the occupation of one of the important d orbitals is different from that of the other. Double exchange was first suggested by Zener (1951) and worked out in some detail by Anderson and Hasegawa (1955). We shall illustrate their discussion with another simple example.

Suppose the magnetic ions concerned are Mn^{3+} and Mn^{4+} . Then it is possible to get a ground state described by occupying the orbitals as follows:



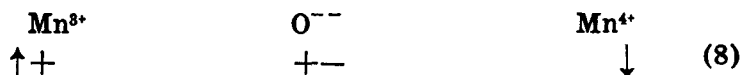
where in this diagram we have used + and -- as before to denote occupation of the important d orbitals and the arrows \uparrow or \downarrow to denote the total spin of the d electrons in the other orbitals. Thus, in this state (6) the important d orbital of the Mn^{3+} ion is occupied

while that of Mn^{4+} is empty, and the total spins are parallel. We notice that if we move the electron $+$ from one magnetic ion to the other we get:



We notice that state (7) has the same energy as state (6) and that there is certainly a matrix element connecting them. Effectively, therefore, an electron can move from one magnetic ion to another preserving its spin orientation. Because the "moving" electron ($+$) is coupled to the other electrons (\uparrow) by intra-atomic exchange, a net coupling between the ions is produced which we can show is always ferromagnetic.

If the matrix element connecting (6) to (7) is b , then the coupling produced by this double exchange process is of order of magnitude b at low temperatures and b^2/kT at high temperatures. The way this comes about may be seen roughly as follows. Given that the states (6) and (7) are initially degenerate and that they have a matrix element, b , connecting them, then the two eigenstates are given by the sum and difference of (6) and (7) and the eigenvalues are b and $-b$ respectively. Using the appropriate Boltzmann factors, we find that the mean energy at temperature T is $-b/\tanh b/kT$. But if the spins were antiparallel, i.e., if we started from:



then double exchange could not take place without violating Hund's rule. It follows from this result that, if the susceptibility χ is expanded in inverse powers of T , then no term of order $1/T^2$ will appear, i.e., the Curie-Weiss constant θ must be zero for double exchange. Notice that in this double exchange process, the final result is to transfer an electron from one magnetic ion to another. We therefore expect double exchange to be accompanied by a high electrical conductivity.

But actually, as Anderson (private communication) himself has pointed out, this theory of double exchange as given by Anderson and Hasegawa (1955) deserves very careful re-examination because it is hard to see how crystal distortions, of either a local or of a cooperative nature, can fail to stabilize either (6) or (7). For example, in (6) we would expect the O^{--} ion to move to the right so as to get closer to the Mn^{4+} ion, but then it costs energy to transfer an electron from the magnetic ion on the left-hand side to that on the right-hand side.

Assuming therefore that this stabilization takes place so that (6) is lowered by an amount, say w , relative to (7), then provided $w \gg b$, an electron can no longer be transferred in a real process. In such circumstances the electrical conductivity must be reduced very considerably, and the real double exchange process must be replaced by a "virtual double exchange" process. This gives rise to a coupling which now may be rigorously described by the Dirac vector model and of order of magnitude b^2/w independent of temperature. It follows that virtual double exchange leads to a non-zero, positive Curie-Weiss constant.

The essential difference between superexchange and virtual double exchange is that in the superexchange processes, electrons are added to the magnetic ions, but in the double exchange process one of the magnetic ions loses electrons. Notice that, if it exists, double exchange, because it gives a coupling of order b which is of second order in the overlaps, may be very much stronger than superexchange or virtual double exchange which give couplings only of fourth order in the overlaps.

Virtual double exchange can also be important in cases where it is clear right from the beginning that no real transitions can take place; for example, $\text{Mn}^{3+} - \text{O}^{2-} - \text{Cr}^{3+}$ with the same orbital occupation as in (6) gives ferromagnetic coupling.

We may summarize all these rules for the sign of the coupling in the following table which lists: n_1 , the occupation number of one important d orbital; n_2 , the occupation number of the other important d orbital, and the signs to be expected from the single transfer superexchange process SE I; the double transfer superexchange process SE II; and the virtual double exchange process VDE:

n_1	n_2	SE I	SE II	VDE
2	2	0	0	0
2	1	0	0	f
2	0	0	0	a.f.
1	1	a.f.	a..	0
1	0	f	f	f
0	0	0	a.f.	0

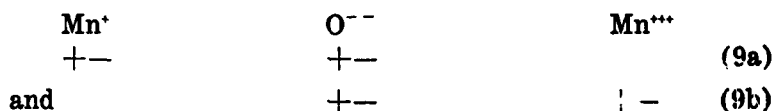
Notice that at no time are different signs predicted by different processes. Hence we may be reasonably confident of these results. It is very dangerous to speculate on the magnitudes of the coupling to be expected in each case, but it seems likely that the coupling will be strong only for the cases (1 1) and (1 0).

The most complete test of these rules available to date is prob-

ably that of Wollan and Koehler (1955) who made a very thorough neutron diffraction study of the perovskite compounds $[(1-x)\text{La}^{3+}, x\text{Ca}^{2+}][(1-x)\text{Mn}^{3+}, x\text{Mn}^{4+}]\text{O}_6^{--}$. By changing composition, i.e., x , it is possible to create crystals with various kinds of nearest neighbor orbital occupations. It is observed that sometimes the coupling is ferromagnetic and sometimes antiferromagnetic, and it is, in fact, possible to correlate all of the unambiguous experimental results with the rules we have just given.

All the discussions we have given so far have been purely qualitative but for actual calculations it is necessary to have a precise formulation of the problem. Recently Keffer and Oguchi (1959) have shown how to obtain this, and it seems likely that calculations along the lines they indicate will prove very rewarding. Another interesting, and quite different, approach to a precise formulation has been given by Anderson (1959B).

We should emphasize once again that we cannot be at all sure that we know all the important ideas governing exchange effects in salts. It is worthwhile to illustrate our ignorance by pointing out that even for the best understood example, $\text{Mn}^{++} \text{O}^{--} \text{Mn}^{++}$, it is possible to think of a new effect, nowhere mentioned in the literature, which it seems likely is reasonably important. This effect is the antiferromagnetic coupling produced, in second-order perturbation theory, by mixing into state (1) the excited states which we can write as:



From estimates, we conclude that we have every right to expect this new process to be more important than the "double transfer" process and possibly to be comparable to the "single transfer" process. Clearly a great deal of theoretical work needs to be done in this subject. It is equally clear that experimental work is desperately needed.

There are several lines of work that seem likely to be profitable. Neutron diffraction investigations of ordering patterns, particularly in crystals whose composition may be varied, would be well worth while. Investigations such as that of Griffiths et al., (Griffiths, Owen, Park and Partridge, 1959) on K_2PtCl_6 doped with Ir give important and direct information on the exchange coupling; these investigations are particularly welcome since they do not suffer from the same uncertainty as other methods which attempt to deduce J from observed transition temperatures using some

approximate theory of cooperative phenomena. Valuable information is obtained by measuring the hyperfine interaction between the electron spin and the nucleus of the anion (Tinkham, 1956; Shulman and Jaccarino, 1956). Although there is no direct relation between this hyperfine interaction and the exchange coupling, both arise from similar processes and hence information on the former can prove an important guide to calculations on the latter.

3.1.2. Dilute Alloys

The study of dilute alloys is becoming increasingly important in helping us to understand the electronic structure and magnetic properties of metals. The alloy system most extensively studied is Mn in Cu (Owen, Browne, Knight and Kittel, 1956; Owen, Browne, Arp and Kip, 1957) and we shall illustrate the ideas involved by considering this alloy as an example. Experiment indicates that in the dilute alloy, each Mn has a spin of either $5/2$ or 2 . High temperature measurements of the magnetic susceptibility give a law $\chi = c/(T - \theta)$ indicating ferromagnetic coupling between the Mn spins but actually all the alloys studied (up to 11 percent Mn) become antiferromagnetic at a Néel temperature T_N which is greater than θ .

Clearly we must consider how we can get both ferromagnetic and antiferromagnetic coupling between Mn atoms which are quite well separated in a dilute alloy. It seems that this relatively long-range coupling must be produced in some way by the conduction electrons, but, at first sight, it appears possible for them only to give ferromagnetic coupling. The simple argument leading to this conclusion has been given by Zener (1951) and by Owen et al., (Owen, Browne, Knight and Kittel, 1956) and essentially is as follows: there is a ferromagnetic exchange interaction between the localized Mn spins and the conduction electron spins. This s - d exchange integral G is approximately 0.2 eV for Mn. Hence a single Mn spin produces a parallel polarization of the uniformly spread conduction electrons. If a second Mn spin is now introduced, it will be coupled parallel to this uniform polarization (and hence parallel to the first Mn spin) by the s - d exchange interaction. The result is an effective ferromagnetic coupling between the Mn spins of the order of magnitude of G^2/E_F , where E_F is the Fermi energy. The error in this simple argument was pointed out by Yosida (1957) who made a careful calculation of the coupling produced. Yosida showed that, in the simple argument, although we use first-order perturbation theory to treat the s - d interaction effects, the net energy change produced is only of second order in

the interaction. This is because the *s-d* exchange interaction lowers the energy of the conduction electrons with spins parallel to the Mn spins but raises the energy of the conduction electrons with spin that are antiparallel. To first order in the interaction, i.e., G , these effects cancel, and the net energy change is therefore only of order G^2 . Because of this rather unusual circumstance, second-order perturbation theory produces energy changes of exactly the same order of magnitude as first-order perturbation theory. Hence, we cannot expect the simple argument to give the correct answer. With two simplifying assumptions:

- (a) the conduction electrons may be treated as free electrons; and
- (b) the radial extent of the localized *3d* electrons on Mn is small compared to the wavelength of the conduction electrons at the Fermi surface,

Yosida found the effective exchange parameter to be given by

$$J_{ij} = \frac{-18\pi n^2 G^2}{E_f} F\{2k_0 R_{ij}\} \quad (3.4)$$

where

$$F(x) = \frac{x \cos x - \sin x}{x^4} \quad (3.5)$$

n being the number of conduction electrons per atom (unity for Cu), k_0 the Fermi wave vector, and R_{ij} the distance between Mn atoms. We notice that the last factor of this expression for J_{ij} contains the only dependence on the distance R_{ij} and oscillates in sign. Hence, some Mn atoms may be coupled ferromagnetically and some antiferromagnetically depending upon their distance apart. Notice that this exchange coupling is of quite long range; at large distances it falls off as $(2k_0 R)^{-3} \cos 2k_0 R$. The oscillations in sign of J_{ij} are determined on a distance scale fixed by the Fermi vector k_0 . The particular form of J_{ij} given here actually predicts antiferromagnetic coupling for two Mn atoms which happen to be nearest neighbors and ferromagnetic coupling for next-nearest neighbors. But we must not rely quantitatively on this last conclusion because the distance scale determining the oscillations in sign of J_{ij} may very well be sensitive to the particular simplifying assumptions, particularly (b), which Yosida made. The important point is that this conduction electron mechanism produces both ferromagnetic and antiferromagnetic interactions, and experiment shows interactions of both signs to be present.

Given the Yosida mechanism, it is still not easy to work out the statistical mechanical properties of the alloy because the Mn atoms

are distributed at random. A good start on this theoretical problem has been made by Sato, Arrott and Kikuchi (1959) who essentially consider the case where the coupling is negligible except for spins which are nearest neighbors. Further experimental and theoretical work on dilute alloy systems would be very interesting.

We should not leave this topic of exchange in dilute alloys without mentioning other possible mechanisms. It has been suggested (Weiss, private communication) that the Mn spins may be coupled by a superexchange interaction via the nominally inert Cu atom. This does not seem likely at first sight but later we shall see that there is fair evidence that superexchange can occur via Au atoms so we certainly must not dismiss the idea. Pratt (1957) has suggested that the Mn spins may be coupled by a superexchange mechanism utilizing conduction band states in place of the anion p orbitals used in salts. Recently Overhauser (1959, 1960) has suggested a new mechanism of exchange coupling in the dilute alloy systems based on the existence of static spin-density waves in the Fermi gas of the metal. This new state of the gas is dynamically self-sustaining as a result of the exchange potential arising from the spin-density distribution. The paramagnetic solute atoms are then oriented by their exchange interaction with the spin-density wave, the resulting energy decrease from this interaction more than compensating for the energy increase associated with the spin-density wave. This theory has certain attractions which may make it applicable to other examples of antiferromagnetism, and may, in fact, also explain the anomalous antiferromagnetic properties of the pure transition elements, e.g., Cr and Mn. Overhauser's theory accounts for the antiferromagnetism observed in dilute Mn-Cu alloys even though there is a positive exchange interaction since above the Néel point the Overhauser interaction vanishes and the only interaction present in the Curie-Weiss behavior is the direct Mn-Mn exchange in which positive interactions could predominate. This theory is also consistent with the recently observed absence of paramagnetic scattering in Cr above the Néel point.

3.1.3. Rare Earth Metals

These metals have complicated magnetic structures, sometimes ferromagnetic and sometimes antiferromagnetic. It seems quite certain that direct exchange is negligible in these metals because the magnetic $4f$ electrons are so deeply buried in the atom that they have negligible overlap with the $4f$ electrons of another atom. It is very likely, indeed, that an indirect exchange mechanism such as the Yosida mechanism or an Overhauser mechanism is

that which couples the spins. The Yosida mechanism gives ferromagnetic coupling at some distances and antiferromagnetic coupling at other distances that would be expected to lead to complicated magnetic structures, just as are observed. Neutron diffraction determinations of the structures with a view to correlating the results with what might be expected with a Yosida mechanism are likely to prove fruitful and interesting.

In performing and interpreting such experiments it should be remembered that the spin pattern in rare earths is not determined entirely by the exchange coupling because of the following reason: for the rare earth metals the spin orbit coupling is so much stronger than crystalline field effects that the orbital magnetic moment is not quenched at all. Hence the spin S and the orbital momentum L couple to give a state of given total angular momentum J , and then the degeneracy between the $2J + 1$ states of given J is split by the crystalline field. This, of course, seriously affects the value of the magnetic moment to be expected from each atom (it will be quite different from the "spin only" value) and also gives rise to strong anisotropic properties (because the unquenched orbit sees the effect of the crystal fields strongly.) One example of this strong anisotropic behavior is given by the very large fields needed to magnetize a rare earth ferromagnetic in a hard direction. These anisotropic effects upon spin orientation can be comparable in magnitude to the exchange effects and thus can influence the spin ordering which takes place.

It should also be remembered that the crystal field splitting of the $2J + 1$ states is quite small and their occupation therefore varies widely with temperature. As different states with different properties become occupied the strength and range of the Yosida coupling will vary and so, too, will the anisotropy effects. We must therefore not be too surprised if the spin pattern and effective exchange couplings vary widely with temperature for these metals. It seems plausible that this is the origin of the ferromagnetic-antiferromagnetic transitions observed in a few cases.

3.1.4. Transition Metals

Our knowledge of the origin of exchange effects in these metals is very poor indeed, and it seems quite certain that a vast amount of experimental and theoretical research remains to be done before we can really begin to understand these metals properly.

The problem is further complicated by the fact that recent experimental work (Weiss and De Marco, 1958; Batterman, 1959) measuring X-ray form factors indicates that the electronic struc-

ture of these metals is much more complicated than we believed it to be previously. Whether the outer electrons are for the most part truly 3d electrons, whether they are more "conduction-like" or whether they are 3d electrons of vastly different radial extent from what is expected, we do not yet know, nor are the experiments or their interpretation conclusive. Until the atomic scattering factor is computed for real electrons in energy bands, one cannot draw any firm conclusions from the experiments. It seems quite certain that we must solve these problems of electronic structure before we can begin to understand exchange mechanisms in transition metals.

Several mechanisms have been postulated to give exchange coupling in transition metals. The first of these is, of course, direct exchange as suggested by Heisenberg (1928). The formula given by Heisenberg for the exchange parameter is

$$J = \iint dr_1 dr_2 A^*(2) B^*(1) \left[\frac{e^2}{r_{AB}} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{A2}} - \frac{e^2}{r_{B1}} \right] A(1) B(2) \quad (3.6)$$

where A and B are orbitals centered on atoms A and B respectively. Usually exchange integrals of this kind are negative, i.e. antiferromagnetic, but Bethe (1933) has given an argument to make it plausible that the parameter be positive for the transition metals (as experiment requires for Fe, Co and Ni). This argument has been questioned (Slater, 1936; Zener and Heikes, 1953) and we still do not know if it is correct. The only calculation made to date is that of Kaplan (1952) who found J to be positive, but this result is inconclusive because the computation was not made at the internuclear distance appropriate to an actual ferromagnet but at the very small distance of $0.25a_0$. Anyway it is quite certain that the simple formula for J given by Heisenberg is incomplete. A more correct result, easily derived from Heisenberg's work, gives an exchange parameter of:

$$J_{ij} = J - CS^2 \quad (3.7)$$

where J is the integral given above, and in the correction term, $-CS^2$, the integral C is the Coulomb interaction between atom A and atom B, i.e.,

$$C = \iint dr_1 dr_2 |B(2)|^2 |A(1)|^2 \left[\frac{e^2}{r_{AB}} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{A2}} - \frac{e^2}{r_{B1}} \right], \quad (3.8)$$

and S is the overlap integral:

$$S = \int dr_1 B(1) A(1) \quad (3.9)$$

This correction term tends to make J_{ij} negative. We do not know how important this correction is.

The second exchange mechanism postulated is that of Zener (1951) who suggests that the $3d$ spins may be coupled via the conduction electrons and the $4s-3d$ exchange integral G . We have discussed Zener's argument earlier (in the section on dilute alloys) and we pointed out that it has to be seriously modified in the way indicated by Yosida (1957). It should also be mentioned that Zener's argument gives the right magnitude for effective interaction only if the polarization of the conduction electrons is computed with exchange among the conduction electrons but no correlation. This is well known to give much too large a polarizability for the alkali metals (and even ferromagnetism for cesium). Thus, Zener's mechanism does not give the right result even for his assumption, and the Yosida effect will reduce it even further. Because this mechanism, when correctly considered, gives rise to both ferromagnetic and antiferromagnetic coupling, it seems plausible that for transition metals it is not the dominant mechanism.

The third mechanism which has been postulated by Slater (1936) and by Van Vleck (1953) is that of "intra-atomic exchange." This intra-atomic exchange results from the fact that the $3d$ electrons can move (quite freely in a band picture but to a lesser extent in a Van Vleck model) and therefore have a nonzero chance of sitting on the same atom. While they are on the same atom the Hund's rule energy couples them parallel and the electrons "remember" this when they move apart. Quantitative calculations on this "intra-atomic" mechanism are difficult; it is easy to obtain an estimate of the exchange integral between two $3d$ electrons on the same atom from spectroscopic data but it is hard to take proper account of correlation effects and thus calculate what is effectively the probability of such an occurrence. When correlation effects are neglected, the intra-atomic exchange leads essentially to the Stoner model or to the more refined Slater (1936) version of it. These models are known to over-favor ferromagnetism, but there might be some hope of modifying them in simple ways, e.g., by the use of screened interactions.

Of course the sign of the exchange parameter in pure metals is known from experiment: Ni, Co and Fe are ferromagnets; Mn and Cr are antiferromagnets. Further important information will be given by the study of the magnetic properties of dilute ferromagnetic and antiferromagnetic alloys. Measurements of the saturation magnetization of ferromagnets as a function of

composition and neutron diffuse scattering from them (Shull and Wilkinson, 1955) would be particularly welcome in giving information both on electronic structure and exchange parameters. Recently Lomer (1960) has shown how the susceptibility measurements on Cr-base alloys made by Lingelbach (1958) may be interpreted to give important information.

Of very great interest are studies of intermetallic compounds, particularly using neutron diffraction techniques. For example, work by Bacon and Street (1958) on MnAu shows that the Mn spins are ordered in a way that is only to be understood in terms of a "superexchange" 180° coupling via the nominally inert Au atoms. Sato (1961) has examined this problem from the point of view of an enhanced dipole-dipole interaction. To conclude, we should emphasize that in this field of work it is particularly valuable to examine many properties, using various techniques, of the same sample.

3.2. Experimental Observation of Exchange Effects

There are many measurements from which the exchange integral, J , can be deduced and we shall now list these measurements and briefly explain their theoretical basis.

3.2.1. Curie or Néel Temperature (I)

This method is the most straightforward and merely requires a knowledge of the Curie or Néel temperature, the crystal structure to give the number of nearest neighbors, and the spin of the magnetic atoms. Statistical mechanics provides calculated values of J/kT_c as a function of the number of nearest neighbors and the spin. The statistical mechanical methods all assume the Dirac vector model (above) and only nearest-neighbor interactions. The three most reliable statistical methods used to date are: (1) the Kramers-Opechowski method, (2) the Bethe-Peierls-Weiss method, and (3) the constant-coupling method. Since the Curie temperature cannot be evaluated exactly, the various methods set about to calculate it through various mathematical approximations, and the interested reader is referred to the following two references for the most recent details and calculations (Brown and Luttinger, 1955; Kasteleijn and Van Kranendonk, 1956R). In the main, the various methods agree to within about 20 percent, being more reliable for the larger spin values and the larger numbers of nearest neighbors.

One source of error in the statistical method is the approximation based on use of only nearest-neighbor interactions. Certain

antiferromagnetic substances appear to have rather significant second-nearest neighbor interactions. Furthermore in some substances, e.g., the rare earths, indirect exchange is probably the only relevant mechanism, and this will certainly fall off with distance more slowly than would be implied by considering only nearest-neighbor interactions.

3.2.2. Total Energy of Magnetization (II)

One can show with the Dirac vector model that the total energy required to completely uncouple the spins in a magnetic material (i.e., the magnetic energy absorbed from $T = 0$ to $T \gg T_c$) is:

$$W = Nz JS^2 = \int_0^{T \gg T_c} C_p(\text{Mag.}) dT \quad (3.10)$$

where N is the number of magnetic atoms in the sample, z is the number of nearest neighbors and S is the spin of the magnetic atom. (The energy of a pair of magnetic atoms is $2JS^2$ and there are $Nz/2$ pairs in a crystal.) To determine experimentally the energy of magnetization, one measures the specific heat and separates the magnetic contribution (generally recognizable as a lambda-shaped anomaly). The details of the separation are given by Hofmann, Paskin, Tauer, and Weiss (1956). In the case of Fe, Ni and Gd, the values of J so derived are in good agreement with the values derived by I (Curie temperature), while in the remaining rare earths other than Gd the agreement is within about 25 percent. There is some uncertainty in correctly separating the magnetic contribution to the specific heat from the nonmagnetic, but with good data over a wide temperature range, this appears feasible to within 10 percent, especially since the entropy

$S = R \ln (2S + 1) = \int_0^{T \gg T_c} (C_p(\text{Mag.})/T) dT$ provides an internal check. For approximately 25 metals and compounds Hofmann et al. (1956) found that J plotted against S yielded smooth monotonically decreasing curves for isomorphic compounds (those with the same crystal structure). The significance of this observation is not understood, and further specific heat work is needed to clarify this point.

3.2.3. Low Temperature Magnetization and Low Temperature Specific Heat (III)

At low temperatures $T < T_c/5$, a quantum mechanical treatment of magnetic materials based on the Dirac vector model leads to a spin wave picture for its thermal excitation. In a classical picture

of a spin wave, one can envision the magnetic spins as a picket fence coupled together with springs. If one end of the fence is held fixed and the other end given an angular displacement we have a standing spin wave. If we let go of this end the resulting motion of the picket fence represents a dynamical spin wave. The quantum mechanical spin wave theory for ferromagnets and ferrimagnets leads to the following expression for the magnetization at any temperature $< T_c/5$:

$$M(T) = M(0) \left[1 - C \left(\frac{T}{J} \right)^{3/2} \right], \quad (3.11)$$

where the constant C is a calculable function of spin and crystal structure. For the spin wave specific heat of ferromagnets and ferrimagnets one obtains:

$$C_v = D \left(\frac{T}{J} \right)^{3/2}, \quad (3.12)$$

where the constant D is also a calculable function of spin and crystal structures. Accurate magnetization measurements at low temperatures have been performed on Fe, Ni, and Gd, and the values of J obtained for Fe and Ni are about twice as great as obtained from I and II above. This is a *major* discrepancy in the theory. On the other hand, the value of J obtained for Gd from low temperature magnetization is in good agreement with values obtained from I and II.

The values of J obtained from low temperature specific heat for both Ni and Gd agree with those obtained in I and II. In the case of Fe the low temperature specific heat is too small a fraction of the total specific heat to make an accurate separation. In magnetite, Kouvel (1956) has used low temperature specific heat data to extract a value of J which, in this case, turns out to be too low.

In the temperature range $\sim T_c/20 < T < T_c/5$, antiferromagnetic spin wave theory predicts a T^3 dependence for specific heat and a T^2 dependence for the sublattice magnetization (This is the magnetization of each of the two or more interpenetrating sublattices of magnetic spins. The total magnetization of an antiferromagnet is zero of course.). In the temperature range 0 to $T_c/20$, antiferromagnetic spin wave theory is complicated by the effects of magnetic anisotropy. Unfortunately, antiferromagnetic specific heat curves do not cover a sufficiently large temperature range for an accurate determination of the magnetic specific heat. To measure the sublattice magnetization one can measure the nuclear resonance of a nonmagnetic atom in an antiferromagnet, such as the proton resonance in $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$. The hydrogen nuclei are

so situated in this crystal as to be nearer either the Cu atoms whose spin is up, or whose spin is down, and the field at the hydrogen nuclei is the resultant of the external field and the effective field from the nearby Cu magnetic spin. As the temperature is raised the Cu spins become misaligned, and the resonance frequency is shifted. The number of such measurements are quite meager, and no specific comparison with theory can be made.

At this point we should mention the measurements of Benedek and Kushida on the sublattice magnetization of MnF_2 with pressure through the hyperfine interaction which reveals the increase of the Curie temperature, and hence J , with pressure (Benedek and Kushida, 1960).

3.2.4. Susceptibility at the Néel Temperature (IV)

It can be shown that the susceptibility at the Curie or Néel temperature is given by

$$\chi = A/(J/k) \quad (3.13)$$

where the constant A is a calculable function of crystal structure and spin. Most recently Smart (1959) has analyzed the susceptibility of several antiferromagnetic compounds. In the case of MnF_2 , the values of J/k are in good agreement with those obtained from I (Curie temperature) and II. In the case of LaFeO_3 , LaCrO_3 , KFeF_3 , and KCoF_3 he gets good agreement with I (No specific heat measurements are available to compare with II).

3.2.5. Domain Wall Thickness (V)

In an unmagnetized ferromagnet or ferrimagnet below the Curie point, there is a domain structure such that within each domain the spins are all aligned, but the direction of alignment varies from one domain to another. The regions between two domains consist of spins which undergo some rotation to adjust smoothly to the spin alignment of the two neighboring domains. This transition region is called a Bloch wall. Since the exchange integral tends to align all atoms parallel, it might be expected that the rotation of magnetic dipoles would occur quite slowly as one passes from one domain to the next. However, the anisotropy energy, i.e., the preference for the dipoles to point in a favorable crystal direction, would tend to make the rotational adjustment more abrupt so as to have as few dipoles as possible in an unfavorable crystal direction (the rotation of spins between two domains aligned 180° apart must indeed pass through many crystal directions). The thickness of the Bloch wall adjusts itself between

these two competing processes, and it can be shown semiclassicaly that this thickness is given by:

$$t = (\pi^2 JS^2 / Ka)^{1/2}, \quad (3.14)$$

where K is the anisotropy constant and a is the lattice constant. (No quantum mechanical treatment of domain walls is available.) In general domain walls are several thousand angstroms in thickness, and attempts to measure the thickness to obtain J have proved to be exceedingly difficult. This has been done by directly measuring the energy stored in a domain wall or by using polarized neutrons of the right velocity such that their natural precessional frequency in the magnetic material enables it just to follow the spin rotation as the neutron passes through the domain wall. Unless some novel experimental technique is devised, this method appears less accurate than methods I-IV.

3.2.6. Inelastic Neutron Scattering (VI)

Just as the inelastic scattering of neutrons (which absorb or excite lattice waves) can be related to the force constants between atoms, so can the inelastic scattering of neutrons (which absorb or excite spin waves) yield information about the exchange integral. This potentially is a most powerful tool for it can determine the magnetic energy levels at all temperatures. Unfortunately, these measurements are still quite difficult, but as higher flux reactors are developed this situation will improve. At temperatures $T < T_c/2$ the energy of a spin wave is:

$$E = AJq^2, \quad (3.15)$$

where A is a calculable constant depending on spin and crystal structure, and q is the wave vector of the spin wave. If one measures the energy and momentum change of the neutron when it is inelastically scattered from the magnetic spins, then one can determine J . Not many experiments have been performed of this type due to experimental difficulties but in one such experiment on Fe_3O_4 the q^2 dependence has been verified.

3.2.7. Paramagnetic, Spin Wave, and Ferromagnetic Resonance (VII)

Recently Griffiths et al. (1957) have replaced some of the Pt atoms (spin 0) in diamagnetic $PtCl_6$ by Ir atoms which have spin $1/2$. The isolated iridium atoms will undergo a resonance at microwave frequencies, and there will be a few isolated cases where iridium atoms will be nearest neighbors. At any finite temperature, some of these Ir pairs will have parallel spins due to

thermal excitation and will resonate as a pair (spin 1) at a different frequency than the isolated atoms (at absolute zero these Ir pairs will be antiparallel with no resultant spin). The energy required to reverse a pair of spins is $E = 4JS^2$. By studying the temperature variation which varies as $e^{-E/kT}$, one can determine E . The value of J so deduced agrees with the value of J deduced from susceptibility measurements at high temperatures on the same sample. These measurements are only feasible when the magnetic pairs are sufficiently far from the other magnetic atoms (several atomic distances) such that magnetic interactions with them are negligible.

In the very thin Permalloy films, Seavey and Tannenwald (1958) have set up standing spin waves pinned at the end. This occurs in the microwave range and provides a direct measurement of J from the frequency of the standing waves, the thickness of the film and the magnetization. This affords great promise but to date the experimental techniques need some refining before a specific value can be quoted. However, this experiment has shown that J itself varies only slightly over a wide temperature range.

A method for measuring J limited to magnetic metals of very low magnetic anisotropy has been used by Rado and Weertman (1954) who observed the broadening of the ferromagnetic resonance near the surface of a metal. Ordinarily the ferromagnetic resonance frequency is simply proportional to the square root of the product of the external field H and the internal field of a ferromagnet B . However, within the skin depth of a metal the resonance is broadened in a complicated fashion related to the exchange integral. Some controversy still exists about this method, but it does give results within a factor of two of that obtained from I above.

3.2.8. Resistivity of Magnetic Materials (VIII)

At temperatures above the Curie point where the magnetic dipoles are disordered, ferromagnetic and antiferromagnetic metals exhibit an additive temperature-independent magnetic resistivity given by:

$$\rho \propto G^2 S(S + 1), \quad (3.16)$$

where G is the strength of interaction between the current carriers and the magnetic electrons. This interaction can be described by the Dirac vector (point interaction) model as:

$$H = 2 \sum GS \cdot s \quad (3.17)$$

where S is the spin of the magnetic atom and s is the spin of the conduction electron. In the case of the rare earth metals, it is believed that the conduction electrons are responsible for the magnetic coupling between atoms since the $4f$ electrons do not extend far enough to interact directly between neighboring atoms. The value of G obtained from resistivity is ≈ 0.05 ev for all the rare earth metals. If these current carriers are responsible for the magnetic coupling, then one expects $G^2/E_F \cong J$. For the rare earths $G^2/E_F = 2.5 \times 10^{-4}$ which agrees with 2.5×10^{-4} ev for J as obtained by I, II and III. In the case of the transition metals $G \cong 0.7$ ev giving $G^2/E_F = 0.05$ ev, to be compared to 0.02 ev for J . The variation of ρ with $S(S + 1)$ appears to be followed quite accurately for both transition metals and rare earths. One would expect the point interaction model to be much better for highly localized f -electrons than for less localized d -electrons. Thus better agreement between values of J computed from resistivity and from Curie temperature *would* be expected for the rare earths. For the transition metals, as suggested by Friedel, one may get a sort of Ramsauer effect arising from a very large phase shift in the d -wave scattering of the conduction electrons. This large d phase shift may contribute much more to the scattering than it does to the interaction, since scattering and interaction depend on phase shifts in different ways. This subject requires more detailed theoretical study since many simplifying assumptions are made to describe quantum mechanically the current carriers in the magnetic metals.

3.2.9. Perpendicular Susceptibility Below the Néel Temperature (IX)

For antiferromagnetic single crystals in which the crystalline anisotropy fixes the spin direction along a unique axis, it is possible to measure the susceptibility perpendicular to this direction. In the molecular field approximation (this neglects short range order effects), this susceptibility is constant from $T = 0$ to $T = T_c$ and is given by:

$$\chi_1 = \frac{g^2 \mu^2}{4Jz},$$

where g is the spectroscopic splitting factor, μ is the Bohr magneton and z the number of nearest neighbors. The next nearest-neighbor interaction does not affect the value of χ_1 in the molecular field approximation, and it is likely a correct calculation would show their effect to be small. Insufficient data are available to study this method systematically.

3.2.10. Antiferromagnetic Resonance (X)

The antiferromagnetic resonant frequency is related to the exchange integral as follows:

$$\nu \propto (zJM)^{1/2}, \quad (3.18)$$

where M is the magnetization. Johnson and Nethercott (1959) obtain values of zJ in good agreement with values obtained by I, II, and IX, and these are given as follows

Method	zJ
Antiferromagnetic Resonance	3.95
Curie Temperature	3.73
Specific Heat	3.82
Perpendicular Susceptibility	4.25

4. STATIC PROPERTIES OF FERROMAGNETIC AND FERRIMAGNETIC MATERIALS

4.1. Intrinsic Properties

4.1.1. Magnetization

The magnetic moments of free atoms can be calculated from the spectroscopic configuration. Spectroscopic configurations are associated with the combinations of the states of the central-field approximation ($1s^2$, $2s^2$, $2p^6$, etc.) using the Russell-Saunders coupling scheme, which applies if spin-orbit coupling is small. Analyses of spectra in terms of Russell-Saunders coupling leads to several generalizations, the most important of which is, perhaps, Hund's rule; namely, that for a given set of central-field states of the individual electrons, the term with the lowest energy is that with maximum multiplicity of spin (and if there is more than one term value with this maximum multiplicity, the lowest is that with the greatest L). For a number of elements, the magnetic moment of the free atom has been observed in the vapor and found to be in agreement with the predicted value. Furthermore, Russell-Saunders coupling has been found to be a good approximation for all cases in which only electrons involving a single azimuthal quantum number are involved.

For certain of the rare earth elements the magnetic moments corresponding to the free atoms have been observed in solutions, certain salts, and even in a metal, e.g., Gd, for which the ground state is 8S . The magnetic moments of the rare earth ions can be predicted from the ground state of the free atoms because the $4f$ electrons are only slightly perturbed by interaction with other atoms.

For the elements of the second and third transition series, evidence suggests that an atomic description of electron states is only fruitful in predicting magnetic moments in solutions or salts if crystal effects, spin-orbit coupling effects, etc., are taken account of in the first order, sometimes even causing a breakdown of Hund's rule. For the metals and alloys of the elements of the second and third transition series, a more appropriate description appears to be in terms of the "Fermi-gas" model of electrons in a box, the periodicity of the lattice being taken into account through the parameters describing the band structure, as for example, the variation of the density of states with energy. It is significant that, with all the elements of the second and third transition series, in only a few combinations is ferromagnetism observed, and these are primarily in combination with elements of the first transition series e.g., Cr-Pt, Ni-Pd, Ni-Pt, Fe-Pd, Fe-Pt. Whether or not some of these elements or combinations thereof will emerge as either weakly ferromagnetic or antiferromagnetic remains to be seen. Much data on magnetic, electrical and thermal properties are needed.

The elements of the first transition series are intermediate with respect to the Fermi-gas model and the strictly atomic description because certain features of both are necessary in order to encompass the phenomena. For solutions and salts, the atomic description is quite useful if account is taken of the effect of crystalline fields in splitting atomic levels. This splitting causes quenching or partial quenching of the orbital angular momentum. Where the orbital moment is completely quenched the magnetic moments observed in solutions and salts can be explained on a "spin only" basis (using Hund's rule). There are exceptions associated with ions (not in S states) when, in crystalline fields of particular symmetries and intensities, the orbital angular momentum is not quenched.

The magnetic moments of the metals and alloys of the first transition series elements are difficult to explain. The essential features are that the moments of the pure elements do not fit any simple atomic model and show variations in alloys which are rather simple and systematic when viewed with respect to electron concentration. These features have been a major impetus to viewing ferromagnetism on a Fermi-gas (collective electron) model. Moments are explained as resulting from partial or complete polarization of electrons in a band which is characterized by a high density of states. The collective-electron model suffers from a lack of structure in real space, whereas there are a considerable

number of alloying and order-disorder effects on magnetic moments which receive a ready explanation on an atomic basis. The individual atoms are assigned characteristic moments (which do not correspond to numbers obtained from a simple atomic picture). A great deal of progress may be in view by generalizing the collective model to include spatially varying spin density. Overhauser's (1959) theory is one version of this, and a similar model could certainly be used to explain many of the apparently local effects in neutron diffraction, order-disorder, etc.

The increasing amount of neutron diffraction data on the magnetic moments associated with each species of atom in alloys and intermetallic compounds is making it more difficult for a simple model to cover the many observations. It is becoming clearer that the local character of the carriers of magnetic moment on those atoms which enter into cooperative magnetic phenomena is closer to the atomic model than the gas model, but at the same time the "conduction" character of electrons plays the dominant role in the exchange mechanisms which lead to cooperative phenomena, either ferro-, ferri, or antiferromagnetism. If a working model is built starting from a suitable linear combination of atomic functions (to obtain the fractional magnetic moments) and including the exchange and correlation terms necessary to make the model satisfactory, the resulting complexity detracts considerably from the value of the observations of magnetic moment per atom in fixing the model. Lost, for instance, would be the simplicity of the collective electron model in using these moments to assign numbers to the relative occupation of *s*-like and *d*-like states.

A question which may well be asked from the point of view of materials perspectives is whether a maximum value of the saturation moment has already been in fact achieved in Fe-Co with approximately $2.5 \mu_B$ per atom. In the transition series, the maximum to be expected from the "spin-only" description so useful in solutions and salts would be five Bohr magnetons per atom. The question of our ability to achieve $5 \mu_B$ is motivated by attempted application of Hund's rule (the maximum multiplicity of spin) to metallic systems. From a quick view of the results for most of the three groups of transition elements either as metals, alloys or intermetallic compounds, it appears that Hund's rule is not generally operating. Therefore, the atomic functions used to build a working model should include configurations which do not obey Hund's rule as well as those which do. Hund's rule appears to work in a number of solutions and salts where the atoms are ionized and mixing of atomic states of *d*- and *s*-like character is

minimized. In metals where the mixing of *s*- and *d*-like character is large, the strict application of Hund's rule appears unjustified. This is consistent with spectrographic data on the elements which show that the energy separation between states of maximum multiplicity and those of lower multiplicity (for a given configuration) is much less when that configuration is an admixture of *s* and *d* than when it is pure *d*. If Hund's rule were applicable, the choice of the approximate combination of atomic functions for describing a metallic system could be guided by observations of magnetic moments, but if Hund's rule is not applicable, then magnetic moment data loses its strong influence on fixing the atomic description. Nevertheless, the variation of magnetization with temperature and pressure and with chemical composition, crystal structure, and degree of order remains a very useful check on the correctness of any model of electronic structure of metals and alloys. More measurements of such variations on well characterized materials would be valuable.

In the rare earth metals, where Hund's rule has validity, much higher moments per atom are obtained, i.e., $7 \mu_B$ for gadolinium, but the atomic volumes are a factor of three larger than in the first transitions series. Thus, the magnetization of Gd is only a few percent higher than that found in Fe-Co. As Gd has a Curie temperature of 16°C , it is only at very low temperature that the magnetization of Gd exceeds the room temperature value for Fe-Co. (The high magnetization in Fe-Co persists until the phase transformation intervenes near 1000°C .)

Pauling first postulated that the *d*-electrons of the transition metals and alloys were of two types. Approximately 6 *d*-electrons would contribute to the bonding of the metals and, therefore, have their spins paired off, leaving only the remainder free to be polarized. This model would limit the saturation magnetization to $2 \mu_B$ plus a possible contribution from conduction electron polarization (presumably $< 0.5 \mu_B$). Some recent evidence supports the qualitative validity of these arguments which, if true, suggest that Fe-Co is already the best that can be achieved in the metallic state. Slater (1937) anticipated a likely maximum of $\approx 2.5 \mu_B$ on the basis of collective electron arguments. The possibility, at least for the present, remains that some new alloy phase or some intermetallic compound could be found which would have a moment of more than $2.5 \mu_B$ per atom. The search for such materials will undoubtedly add much to the clarification of present difficulties in understanding transition metals, their alloys and their compounds.

The magnetization of the salts of metal ions is in general well

understood in terms of spectroscopic configurations and crystalline field effects. The general effect is an antiferromagnetic interaction which gives parallel alignment of spins on sublattices while directing the sublattice magnetizations in opposition to each other. Where a population difference or moment per atom difference between the sublattices exists, there will be a net magnetization. One of the most favorable cases for obtaining a large magnetization appears in magnetite (Fe_3O_4). In its spinel structure half of the trivalent Fe ions occupy the tetrahedrally coordinated sites (A sites), and the remainder of the trivalent and all of the divalent Fe ions are on the octahedral sites (B sites). As there are twice as many B sites as opposing A sites, there is a net moment due to the uncompensated divalent ions. This is $4 \mu_B$ for each group of three iron ions. As the iron ions are a bit less close together in magnetite than in metallic iron, the magnetization is slightly less than half that found in pure iron. By substituting Mn^{++} for Fe^{++} the net moment is raised from $4 \mu_B$ to $5 \mu_B$ per trio of cations ($5 \mu_B$ is never quite achieved in practice because the Mn ions do not all go on the B sites).

Magnetite is an inverted spinel. In a normal spinel all the B sites are trivalent and the A sites divalent. In the normal spinel it would be possible in principle to increase the moment to $10 - 1 = 9 \mu_B$ per trio of cations and thus almost match the magnetization of iron. (Such a material would be expected to have a Curie temperature below room temperature).

In structures other than spinel different situations could occur. It has been pointed out that for every antiferromagnetic material (of which there are a large number) it should be possible by proper substitution to form a ferrimagnet.

To achieve a higher magnetization than is found in the spinels it would be necessary to increase the 2 to 1 population ratio of opposing sublattices. No presently known antiferromagnetic material accomplishes this.

Because saturation magnetization is so directly related to the valence and the crystal structure, its study is basic in any investigation of salts with new structures or with different chemical compositions. In the spinels, for instance, magnetization measurements are particularly useful in revealing deviations from stoichiometry or the degree to which the structure is normal or inverted.

4.1.2. Experimental Determination of Spontaneous Magnetization

We shall now describe those experimental techniques which give

information about the arrangement and magnitude of spins in a magnetic material as a function of temperature.

a. The only measurement capable of giving a precise and unambiguous description of the arrangement and magnitude of spins on the crystal lattice is neutron diffraction. To date the bulk of neutron diffraction work has been devoted to determining the arrangement, magnitude and spin direction in the long-range order region below the critical temperature. For these cases the technique has been most successful. The variation of long-range order with temperature follows approximately the Brillouin-shaped curves observed for the magnetization of ferromagnetic materials. Some differences have been noticed in the Brillouin shape for some antiferromagnetic compounds but no detailed study has as yet been made. In virtually all cases the long-range order arrangement of spins are colinear (i.e., all spins are either up or down) although the possibility of angles between spins and spiral arrangements has been suggested and sought (Yafet and Kittel, 1952; Kaplan, 1959). In the case of Fe the short range ordering of spins, especially near the Curie point, has been studied in several laboratories both experimentally and theoretically, and in general good agreement has been found. No detailed study has as yet been made of the short-range order below the Curie point. No theoretical or experimental work has solved the problem of the time-dependent correlation of spins, i.e., precisely how spins on various lattice sites change their orientation with time at various temperatures.

b. Another method for measuring long-range order is to measure the net magnetization. In the case of ferromagnets or ferrimagnets the experimental problems in such a measurement have been quite accurately studied. At temperatures close to the Curie point, the external field will induce some magnetization, and one must measure the magnetization as a function of external field and extrapolate to zero field. For pure metals and for ferrimagnetic compounds of proper stoichiometry quite good results can be obtained. In the case of ferromagnetic alloys, though, the lack of homogeneity causes the magnetization curves to be distorted when the Curie temperature varies with alloy concentration. Even for a theoretically homogeneous alloy, the finite number of nearest neighbors must give rise to local variations in the fraction of nearest neighbors of each species, i.e., a 70-30 bcc alloy (eight nearest neighbors) while consisting on the average of 5.6 and 2.4 nearest neighbors of each species will be comprised of local configurations of 6 and 2 and 5 and 3. To the extent that ferromagnetic coupling is

only a nearest-neighbor interaction, the strength of the local coupling will be affected by these fluctuations. This effect requires more study. In the case of rare earth metals, mention should be made of the large anisotropies that can be present and, as such, single crystals may be required to accurately determine the saturation magnetization.

Another technique for measuring spontaneous magnetization applicable to certain antiferromagnets is to measure the nuclear resonance frequency of a nonmagnetic atom whose position in the lattice places it closer to the magnetic atoms with spin up or spin down. The resonant frequency will then depend on the external field and the internal field due to the nearby magnetic atom. This internal field is proportional to the magnetization or long-range order of these particular atoms, and the temperature variation of the resonant frequency will measure this magnetization. This has been done for the fluorine nuclear resonance in MnF_2 , FeF_2 , and CoF_2 and for the proton resonance in $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$.

c. By measuring the specific heat of a magnetic material over a wide temperature range, it is possible to separate that portion of the specific heat which is absorbed by the magnetic spin system. The entropy can then be calculated as a function of the temperature and be related to the spin arrangement at any temperature by:

$$S = R \ln W,$$

where W is the number of ways one can arrange the spin system in a mole of the material, without producing the same configuration, divided by Avogadro's number. When the spins are completely uncoupled, the entropy is $S = R \ln (2S + 1)$, where S is the spin of the magnetic atom. (For materials with two types of magnetic atoms the entropy is additive). The specific heat for a wide variety of magnetic substances reveals the presence of short-range order to temperatures from 1.25 times to twice the Curie temperature. To date no attempt has been made to correlate the entropy, as determined from specific heat, and the spin arrangement, as determined from neutron diffraction, although it should presently be feasible for Fe.

At the Curie point the specific heat exhibits a sharp lambda-type anomaly for pure elements and compounds of proper stoichiometry. However, in the case of alloys, the effect of inhomogeneity mentioned (4.1.2.b.) causes the lambda anomaly to become more diffuse. This has only been measured for a few Cu-Ni alloys and requires more study in order to relate it to local inhomogeneities.

d. A measurement extremely sensitive to the degree of order of

a lattice is the resistivity, but quantitative interpretation is extremely difficult due to our inability to describe correctly the current carriers in quantum-mechanical language. While the portion of the resistivity due to magnetic scattering of the current carriers increases qualitatively with the decrease in long-range order, no quantitative comparison is presently feasible. The problem of separating the magnetic resistivity from the nonmagnetic presents some difficulties when the nonmagnetic high temperature resistivity is not linear in T . If it is linear in T (as it generally is) then the separation can be made to within an error of 10–20 percent. To help interpret resistivity data, more extensive transport property measurements are required in magnetic materials. These include (1) magnetoresistance as a function of crystal direction, alloy concentration, applied field direction and temperature, (2) the orientation effect in ferromagnetic metals (i.e., the change in resistivity when the sample is magnetized parallel and perpendicular to the current direction) as a function of crystal direction, alloy concentration and temperature, (3) the normal and anomalous Hall effects, particularly in the rare earth metals.

e. A new phenomenon observed with neutrons is the presence of antiphase domains in Cr metal. Cr is antiferromagnetic, and the ordered antiferromagnetic arrangement appears to make a mistake approximately every 10 atoms although the precise arrangement has not as yet been solved. The Néel temperature and the sharpness and possible splitting of the diffraction lines depends critically on the state and previous history of the sample—varying from single crystal to powder and sometimes differing in two single crystals or two powders of different origin. Kaplan (1959) has suggested that some of the unusual properties of Cr may be due to spiral spin arrangements which could result if both the nearest and next-nearest neighbor interactions are antiferromagnetic. This suggestion, while interesting and worthy of study, does not appear to be completely consistent with the recent neutron diffraction measurements. Another interesting suggestion is that of Overhauser and Arrott (1960), in which the possibility of spin-density waves similar to those suggested to explain the antiferromagnetic properties of dilute alloys of manganese in copper (Overhauser, 1959) is considered, and the existence of antiphase domains in the particular instances observed is justified. The situation in Cr is, indeed, quite complicated and requires further study. A great deal of neutron work on this subject is in progress, and a more complete understanding of the effects should come out of this work.

f. A most useful technique to study the degree to which an external field has saturated a ferromagnetic sample (i.e., aligned all the domains into a single domain) is to measure the depolarization of a polarized neutron beam which traverses the sample. This technique is particularly useful to measure small deviations from complete saturation. Hughes, Wallace and Holtzman (1948) have studied the approach to saturation in Armco iron and find a $1/H$ dependence in fields of 3,000 to 12,000 gauss, falling to lower values in the exponent of H at higher fields. Danan (1959) has also studied this effect in iron by making very careful saturation magnetization measurements and finds agreement with the neutron work as well as the theory of Holstein and Primakoff (1940).

g. A measurement suitable to studying the homogeneity of alloys is the initial permeability as a function of temperature. For ferromagnetic elements and compounds of good stoichiometry, the initial susceptibility slowly rises with temperature, reaching a peak at the Curie point and abruptly dropping as one increases the temperature slightly above the Curie point. For alloys, though, this abrupt drop is "washed out" over a wider temperature range due presumably to local statistical variations of Curie temperature associated with composition variations. No attempt has been made to correlate the "washing out" of the initial permeability at the Curie point with inhomogeneities in alloys. The initial permeability has also been used to reveal the temperature at which crystalline anisotropy changes direction in a crystal.

Metals—Many transition metals are available in a state of reasonably high purity in regard to metallic constituents but some uncertainty exists in regard to interstitial elements such as O, N, and H, since only one technique for analysis is extensively used. This is vacuum fusion and relies on some chemical reaction with the metal to completely liberate the interstitial atoms for analysis. Until some independent method is developed and studied as a check on the vacuum fusion method one may nearly always expect of the order of 0.01 percent or more interstitial impurities. It is also a good idea to inspect metallographically the pure metals since this will often reveal the presence of other phases such as oxides, etc., which in turn will bear upon any claim to low interstitial contamination. The interstitial atom contamination for rare earths is also quite high.

All bulk metallic samples have some preferred orientation which develops in the process of solidification from the melt and this is undesirable for neutron diffraction intensity work (powder method). 200 mesh (or finer) metallic powders, however, appear

devoid of preferred orientation even when sintered and a minimum of one mole of sample is generally required. To prepare alloys for powder neutron diffraction work, sintering should be employed wherever possible. In this method the constituents are weighed, intimately mixed in a ball mill, pressed into blocks in a die and then sintered well below the melting point in *vacuo* or in a non-reacting atmosphere. This method insures proper stoichiometry since the sample can be weighed before and after sintering. If one intends to make neutron measurements on magnetically saturated samples, the disadvantage of sintered samples is that they require higher magnetic fields to saturate (due to the particle size).

If the element or alloy is brittle, such as Mn or σ phase then resistivity measurements are unreliable since fine microscopic cracks may be present. In fact the problems in preparing metallic samples are so varied that the best advice that we can offer is that physicists so engaged should seek advice from metallurgists. The impression of the physics literature is that such cooperation has unfortunately been difficult to establish.

Nonmetals—The major problem in the preparation of nonmetals is the question of stoichiometry. This has been particularly true of the ferrites and, as is the case of metals, the question of preparation has always been an empirical problem. Techniques for analysis for example, of both transition metal content and oxygen content are not well established, and lack of stoichiometry appears to affect the spin arrangements seriously in many cases. For example, in the case of zinc ferrite the specific heat anomaly and the magnetic neutron diffraction peaks are broadened due to small deviations from stoichiometry. In general, it is suggested that the sharpness of the specific heat peak is one of the most sensitive guides to stoichiometry. Here again it is suggested that physicists work closely with physical chemists in the preparation of the non-metallic magnetic materials.

4.1.3. Magnetocrystalline and Magnetoelastic Effects

The exchange interaction does not reflect the symmetry of a crystal lattice. In ordered magnetic structures however, the magnetic moments prefer to lie along certain crystallographic directions and, in addition, this preference can be changed qualitatively, as well as quantitatively by imposition of external stresses. The energy associated with these magnetocrystalline anisotropy and magnetoelastic effects is usually only a tenth of a percent or so of the exchange interaction, although in some materials, particularly

the rare earths (Behrendt, Legvold and Spedding, 1958), it may be comparable with the exchange interaction. Small though they are, these effects are extremely important to the properties of magnetic materials (Kittel and Galt, 1956R).

Although it is possible to gain some information about anisotropy in ferro- and ferrimagnets from the approach to saturation in polycrystalline materials (Gans, 1932; Néel, 1948), reliable measurements require single crystals. The most direct determination is made by measuring the torque required to rotate a symmetric crystal in a uniform field that is strong enough to insure that the magnetization is closely colinear with the field (Shenker, 1959R). This torque (about a sufficient number of axes) can be integrated to give a surface representing the changes of free energy as the magnetization rotates relative to the crystal axes. There are no severe experimental difficulties except in the case of very high anisotropies, in which case corrections to the measurements must be made for the failure of the magnetization to be essentially colinear with the field. The second principal method is to note the alteration in the field required for ferromagnetic resonance as a crystal is rotated (see sec. 5.1, Ferromagnetic Resonance). This method, however, does not give the same information as the first. At absolute zero, the torque method gives directly the dependence of the ground state energy as a function of direction of magnetization, while the resonance measurement gives the alteration of the separation between the ground state and a first excited state. These two measurements are usually in accord. At elevated temperatures, however, the two methods give essentially different averages over the excited states (e.g., Livingston and Bean, 1959). There is, in addition, the possibility that relaxation processes may take place on a time scale that is intermediate between the long time required for a torque measurement and the short time consistent with the high frequencies employed for resonance (Yager, Galt and Merritt, 1955).

The energy surface found by experiment is usually quite simple. For instance, in most cubic materials an expression of the form $E(\alpha_1, \alpha_2, \alpha_3) = K_1 (\alpha_1^2 \alpha_2^2 + \alpha_2^2 \alpha_3^2 + \alpha_3^2 \alpha_1^2) + K_2 (\alpha_1^2 \alpha_2^2 \alpha_3^2)$, where the α 's are direction cosines of the magnetization relative to the cubic crystal axes, suffices for description of the anisotropy. In most cases, K_2 is less than K_1 . There is at least one case of impure garnets (Dillon, 1958; Dillon and Neilsen, 1959) where the second derivative of energy versus angle of magnetization has a sharp cusp at low temperatures. Kittel (1959) has given a probable

explanation in terms of the crossing of energy levels with angle of a hypothecated rare earth impurity ion that is simultaneously in a crystalline electric field and an exchange field making an angle with the crystalline field.

Neither the torque nor resonance methods can be used to measure directly the anisotropy of antiferromagnets, because of their lack of a net moment. The most direct measurements (critical field for flipping of moment directions, field-dependence of Néel temperature, and antiferromagnetic resonance absorption) all measure, at best, a geometric mean of exchange energy. Some information can be obtained from measurements of anisotropic susceptibility above the Néel temperature. This area has been reviewed by Nagamiya, Yosida and Kubo (1955R).

One may proceed some distance with completely phenomenological theories of anisotropy. First of all, as noted above, symmetry relations may be employed to find the proper functions to employ for correlating experimental data. A next step is to assume that the anisotropy is a function of pairwise interactions between neighboring atoms (Néel, 1954). If this were true, all anisotropies could be expressed in coefficients appropriate to various atomic combinations. The most fundamental question, however, is whether the anisotropy is due mainly to the "simple" electrostatic environment of a magnetic ion or if the presence of a neighboring moment is the determining factor (e.g., Van Vleck, 1959). The crystal field model has had some success in predicting and correlating the anisotropies of ferrites (Folen and Rado, 1958; Yosida and Tachiki, 1957), garnets (Wolf, 1957), fluorides (Moriya, et al., 1956), and of isolated ions (e.g., Bleaney and Stevens, 1953R). The only experiments to test this point in metals suggest that the anisotropy of f.c.c. cobalt has its origin also in its electrostatic environment (Bean, Livingston, and Rodbell, 1959). Very little has been done on the theory of anisotropy in metals since the classical work of Brooks (1940). The temperature dependence of the anisotropy has been a popular theoretical topic (Van Vleck, 1959). While Folen and Rado (1958) and Yosida and Tachiki (1957) have had some success with ferrites, recent results on iron (Graham, 1958) appear to conflict with the generally agreed prediction of anisotropy varying with the tenth power of the spontaneous magnetization.

Some other topics that have not been discussed include the phenomenon of magnetic annealing (Graham, 1959) and surface anisotropy (Néel, 1954). They are related in that they both require noncubic environments of magnetic atoms for their existence. In the case of surface anisotropy, this lack of cubic environment is

provided by the surface itself, while possibilities for magnetic annealing include short-range directional order of substitutional atoms, shaped second phase, aligned interstitials, or possibly oxygen in metals. Many critical experiments have yet to be performed in these areas.

The practical implications of magnetic anisotropy are manifested in two directions. First of all, low anisotropies are useful for materials for power and communications applications since permeability and coercive force depend primarily on the ease of domain boundary motion which is related to the anisotropy. By the same token a high anisotropy is desirable for permanent magnets (e.g., Kittel and Galt, 1956R) and some high frequency applications (Jonker, Wijn, and Braun, 1956). In the future, one of the more important uses may be the employment of highly anisotropic materials as generators, controllers, and detectors of millimeter and submillimeter radiation, again because of the critical dependence of resonance frequency on anisotropy.

Magnetoelastic effects in single crystals have not been investigated as thoroughly as anisotropy. The spontaneous magnetostriction in ferromagnetic materials is small ($<10^{-3}$). Most investigators use the resistance strain gauge (Goldman, 1947, 1951) to measure these small distortions that accompany rotation of magnetization. The theory of magnetoelastic effects has not gone far beyond the simple relation of magnetostriction through the elastic constants to the strain dependence of the anisotropy. Some difficulties with this relation have been pointed out by Brown (1953).

The magnetoelastic effects have led to rather wide usage of magnetic materials as transducers, resonators, delay lines and even as high mechanical damping materials for steam turbine buckets (Cochharlt, 1959).

In addition to further experimental work, the future would appear to hold promise for the extension of crystal field theory into the area of magnetoelastic effects. Already, a somewhat successful attempt has been made by Kanamori (1957) to explain the large ($>1\%$) spontaneous distortions in simple antiferromagnets and a treatment of magnetostriction in ferrites has been given by Tsuya (1958).

4.2. Structure-Sensitive Properties

The details of the technical magnetization curve (e.g., hysteresis loss, coercivity, remanence, initial and maximum permeability) are highly structure-sensitive. They have been explained qualitatively,

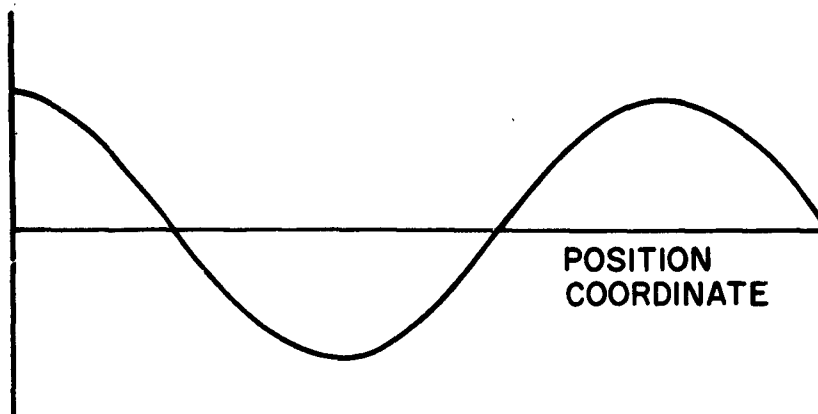
and to some extent quantitatively, by use of the "domain" and "domain wall" concepts and of the "rotation" and "wall displacement" mechanisms. The reader of this section will be assumed to have some acquaintance with these concepts and mechanisms, as they are presented in Kittel's review (1949R), its revision by Kittel and Galt (1956R), or the book by Stewart (1954R). The concepts will therefore not be reviewed in detail; rather, their theoretical and experimental basis will be inspected critically, some shortcomings of the present theory will be pointed out, and possible methods of improving it will be considered.

4.2.1. Domains and Domain Walls

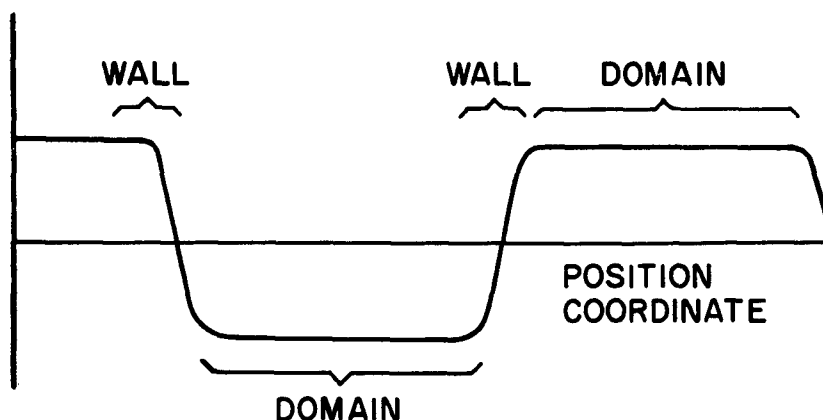
The Domain Concept—(Becker and Doring, 1939R; Kittel, 1949R; Stoner, 1950R; Bozorth, 1951R.) Domains were invented to reconcile the existence of a demagnetized state with the spontaneous magnetization predicted by the Weiss theory. Experimental confirmation seemed to have come with the Barkhausen effect, though the first domain interpretation of this phenomenon was quite different from the present one. The Sixtus-Tonks observation of large nuclei of reversed magnetization provided at least supporting evidence for domains. Finally, more direct evidence was found in Bitter's (1939R) powder patterns, supplemented recently by magneto-optical and electron-microscopic methods. Direct as these may be, they show only the surface structure; with the discovery of transparent ferromagnetics, it has now become possible to get some information about the internal domain structure by optical methods (Dillon, 1958).

The theory of domains has two branches: attempts to explain their origin, and attempts to interpret magnetic behavior in terms of them. Theories of their origin are based on a competition between exchange forces on the one hand and dipolar and anisotropy forces on the other; the former tend to align spins parallel and are short-range, the latter tend (at least to a certain approximation) to avoid poles and are long-range. The expected outcome of this competition is a magnetization direction that varies only slowly with position, yet fast enough to eliminate some or all of the pole field. The crystalline anisotropy forces now modify the situation by pulling the magnetization closer to the "directions of easy magnetization"; thus, instead of a slow variation of direction from one point to another, as in (a), we get a rapid variation over the thin transition layer between one region and the next, as in (b).

(a) Component of Magnetization



(b) Component of Magnetization



Domain interpretations of magnetic behavior began with crude models, each drastically simplified, with no great consistency between one model and another. Considerable empirical success was achieved by a "statistical" theory with no very specific model (Becker and Doring, 1939R; Bozorth, 1951R; Grimes, 1957). But detailed developments required the concept of the energy-laden interdomain wall.

The Wall Concept—(Bloch, 1932; Landau and Lifshitz, 1935). The wall model now in use is essentially that of Landau and Lifshitz (1935). Viewed on a small scale, it is a region of rapid transition, over which the micromagnetization preserves its magnitude but changes its direction. Viewed on a larger scale, it is

a discontinuity surface endowed with surface energy; and because this energy may vary with wall position, there may be forces that tend to move the wall. The Sixtus-Tonks experiments gave indirect evidence in support of this idea; later, Elmore (1938), by the colloid-pattern technique, directly observed such wall displacements. Finally, Néel (1944), by detailed calculations based on the Landau-Lifshitz model and on the principle of pole-avoidance, successfully predicted the details of magnetic structure in picture-frame crystals (Bozorth, 1951R). Since then, the plane domain wall has usually been assumed to be the essential factor in all cases of nonuniform magnetization.

Single-Domain Particles—The Weiss theory predicted uniform magnetization throughout the specimen; there might therefore have been an early search for conditions under which this would in fact occur. Such conditions would have been found in fine particles, in which the long-range magnetic forces do not have room to operate. But instead of this, domains were invented; and once invented, they seemed so inevitable that bodies not possessing them were discovered only by special experimental and theoretical acts (Kittel, 1946; Néel, 1947; Stoner and Wohlfarth, 1948; Guillaud, 1949). Logically, these "single-domain particles" are the normal consequence of the Weiss molecular field; it is the structure of massive materials that should be regarded as anomalous. Once the comparative simplicity of fine-particle magnetism was recognized, it achieved a number of practical successes, both in the interpretation of materials already known and in the production of new ones. The theory developed through several stages and then remained essentially unchanged for some years—capable of predicting the right order of magnitude but not the right values.

Conceptual Situation—In retrospect, it appears that the domain and wall concepts needed some time to win full acceptance, eventually became well established by experiment, and were provided with more or less satisfactory theoretical rationalizations; and, were finally so completely accepted that important alternatives were overlooked. Specifically: (1) the possibility of a uniform non-domain-type structure was investigated only after a long delay; (2) the possibility of a gradual type of spatial variation (to be expected theoretically when the anisotropy is small) has been almost ignored; (3) the wall-displacement mechanism has been assumed to apply in all low-field processes; it may be that other important possibilities are being overlooked.

Theoretical Situation—The theory of the behavior of single-domain particles, given that they are single-domain, is rigorous

for the models assumed; but the models are oversimplified and incapable of producing precise agreement with experiment. The origin of domains is not in a satisfactory theoretical situation; current theories merely compare the energies of alternative models, ignoring the energy barriers between different stable states. The reasoning about domain walls is even less satisfactory. This shows up in the topics to be discussed next.

4.2.2. The Magnetization Process

Mechanisms of Magnetization—Understanding of the magnetization process began with the successful calculation, by Akulov (1931) and others, of the high-field magnetization curves of crystals (Bozorth, 1951R). The model used was a simple one: a uniform vector magnetization capable of rotation, at constant magnitude J_s , in response to torques. The low-field process was recognized as being more complex, was vaguely described as inversion, and was not understood. Becker applied the rotation concept to internally stressed materials, with some empirical success (Becker and Doring, 1939R). Eventually the wall-displacement model became popular; actually, however, Becker's nonuniform rotation and its successor, wall-displacement, differ only because of artificial simplifications present in each. Arguments about rotation vs. wall-displacement (Rado, Folen, and Emerson, 1956) need to be tempered with the realization that sometimes neither of these models is more than a crude approximation.

The Rotation Model—The rotation model of Akulov (1931) worked well for crystals at applied fields H_a such that $H_a J_s$ is comparable with the crystalline anisotropy constant K . The model later found further application in single-domain particles, where the low-field complications seemed to be absent. It could also be taken over intact into many problems of ferromagnetic resonance. Demagnetizing fields, sometimes ignored in the early theory, were ultimately recognized as being important (Kittel, 1947); their calculation in general requires solution of Poisson's equation, and it is only in special cases that "demagnetizing factors" are adequate. Becker's rotation model for material with internal stresses is now practically forgotten, but in cases of low crystal anisotropy it might still be useful.

The Wall Model—In an attempt to calculate initial permeability and coercivity, this model underwent early development in the hands of Becker, Kersten, and Kondorsky (Becker and Doring, 1939R). The wall was pictured as plane; it had a surface energy density that varied if the wall was displaced along the direction

of its normal. Displacement of the wall was understood to proceed by rotation of spins in the transition layer within it, but the details of this process were ignored. Hysteresis was attributed to Barkhausen jumps in which the wall moved from a position newly unstable to one still stable. The factor held responsible for the spatial variation of wall energy was "internal stresses." For these, a very vague model was used; it was years before a dislocation interpretation was attempted (Vicena, 1955).

The "internal stress" era ended when Kersten (1943) proposed a new theory, based on cavities and inclusions as a determinant of wall energy. Soon thereafter, Néel pointed out that Kersten had ignored magnetic energies many times as large as the energies he had taken into account, and Néel (1946) himself published a theory to supersede Kersten's. Kersten's mathematical methods are crude but straightforward; Néel's calculation makes generous use of Fourier series and of ingenious (if somewhat mystical) artifices for applying them to nonlinear problems. Most subsequent calculations have differed from these only in detail.

These theories as a group have certain defects in common. They all postulate walls, assign them simple properties (such as planeness), assume without proof that they retain their character after they become unstable, and concentrate on wall propagation, with only occasional attention to the equally relevant problem of wall nucleation (Goodenough, 1954).

The Situation Regarding Fine Particles—The calculations for the isolated single-domain ellipsoid (Stoner and Wohlfarth, 1948) and for the chain of spheres (Jacobs and Bean, 1955) have the almost unique property, in this field, of being rigorous for the model assumed. It is the model itself that is inadequate. The particle with uniaxial anisotropy has been overworked for lack of quantitative results on other models; the cubic and other cases need to be examined (Tonge and Wohlfarth, 1958; Johnson and Brown, 1959). Magnetic interactions between particles have not been calculated rigorously; Néel's (1947) formula, often quoted as if it were exact, is, in fact, an approximation of unknown reliability. There has been confusion between two meanings of "single-domain"; a particle may be uniformly magnetized in each stable state, yet become nonuniformly magnetized during transitions between such states, with resulting properties different from those of the standard model. The effect of imperfections in destabilizing the uniform state, and the behavior of particles small enough to be destabilized by thermal agitation, need more detailed study. Nevertheless, this aspect of magnetization processes is

generally in much better shape than those related to ordinary materials.

The Situation Regarding Bulk Materials—The wall-displacement model is useful for order-of-magnitude estimates, especially when these can be obtained by simple and quick calculation methods. The model is at best a highly simplified one, and therefore results obtained by applying Fourier series to it do not necessarily have the precision usually associated with that mathematical technique. The empirical successes of the model are not sure evidence of its correctness. There were empirical successes also for the Becker rotation model and for the Heisenberg statistical model; any model with the right energies and symmetries in it has a good chance of hitting the right order of magnitude. There has been too uncritical an acceptance of walls as always existing, as maintaining their identity during jumps, and as being conveniently able to nucleate on demand. There has also been over-concentration on the plane "wall" as against other possible types of magnetic inhomogeneity; the theorist's structures are one-dimensional, possibly because the theorist can solve his equations only in one dimension. It must be remembered, too, that the behavior of an almost perfect picture-frame crystal may differ from the behavior of irregularly stressed, imperfection-filled, polycrystalline bulk material. And when magnetostrictive forces are included, attention should be paid to their shape-dependence (McKeehan, 1950)—a property that tripped up Becker and Akulov but was later cleared up by Powell and Hyasi.

4.2.3. Micromagnetics

Conceptual Basis and History—Landau and Lifshitz treated the inside of a "wall" as a region where the spontaneous magnetization retained its magnitude but changed its direction from point to point. The Landau-Lifshitz wall is a continuum approximation to a lattice: instead of spins at lattice sites, it assumes a vector magnetization whose direction varies continuously with one coordinate. Why not, more generally, assume one whose direction varies continuously with all three coordinates, and thus derive the microscopic magnetization distribution directly, instead of talking about domains and walls? This idea was investigated in a fragmentary way soon after Landau and Lifshitz; but the only solvable problems found were the one-dimensional ones that led back essentially to the wall model (Bitter, 1937R, Shirobokov, 1939, 1945), and linearizable ones that were encountered near saturation (Brown, 1940, 1941). It was the single-domain

particle problem that first led to a more ambitious calculation by this method—a calculation that yielded a rigorous criterion for deciding whether a particle would reverse its magnetization by uniform rotation or by a more complex process, involving at least instantaneously nonuniform magnetization (Kondorsky, 1952; Brown, 1957; Frei, Shtrikman, and Treves, 1957; Arrott and Goldman, 1959R).

Accomplishments So Far—A rigorous solution of this sort has been found only for a perfect crystalline ellipsoid; and the rigorous calculation gives only the value of the characteristic or “nucleation” field at which an initial uniform state becomes unstable. For a body of arbitrary shape or for one containing imperfections, the instability may occur after the deviations from uniformity have become too large to treat by a linear approximation; and after instability occurs, nonlinear equations must be solved in any case. The nonuniform process of “magnetization curling” has been followed through in detail, by numerical integration, for an infinite cylinder, but since longitudinal magnetization of an infinite cylinder involves no poles, such magnetization is always stable and exists, apparently, except transiently during a reversal (Aharoni and Shtrikman, 1958). There is therefore still no rigorous calculation of a stable domain structure. Extension of the theory to a sphere or the like is needed. Approximate calculations by the Ritz method have proved useful, but even they often prove laborious.

Relation to Dynamic Problems—Addition of dynamic terms to the equilibrium equations of micromagnetics (Brown, 1959) gives the equations needed for a complete treatment of ferromagnetic resonance problems in the continuum approximation, and of the “semiclassical” theory of spin waves. Actually, the theory of resonance has been developed step by step, without regard to related equilibrium problems—by successive introduction of the applied field, the demagnetizing field, anisotropy, nonuniform magnetization, and exchange torques. Treatments of switching in memory cores have also made little use of related concepts already developed in static theory. Unification of these several fields is much needed. The “nucleation” fields of the static theory are the zero-frequency limits of the resonance fields, and by going to complex frequencies we should be able to get a linearized theory of the beginning of the switching process; but, the details of these relations still remain to be studied.

Possibilities for Further Applications—The linear instability calculations are by no means complete. They require solution of

transcendental equations and therefore are facilitated by automatic computers. The instability problem needs investigation for a variety of specimen shapes, field conditions, and anisotropy properties, and with imperfections taken into account. Systematic study of resonance and switching problems should be fruitful. Numerical integration of the nonlinear equations, even with computers, is laborious; it must therefore be limited to important and not too complex cases. Such calculations should be done in typical cases, including dynamic ones. Approximate methods, worked out systematically on the basis of dynamic generalizations of the Ritz method, may prove valuable.

Improvements Needed—At the present stage, further refinements of the basic model would be premature. But eventually, the following matters should be looked into, within the framework of a classical continuum formulation. The magnetic energy should be examined critically; the present approximation, based on a Lorentz field, is open to criticism. Surface anisotropy terms may be needed in the boundary conditions. The exchange term needs modification for noncubic crystal symmetries and for materials with two or more magnetic sublattices. Magnetostrictive energy terms should be added and should include the necessary shape-dependent surface terms.

Still further refinement is necessary if one wishes to take into account the finite lattice spacing and the quantum nature of spin interactions. Here the theory begins to run into the fundamental difficulties of the quantum theory of spontaneous magnetization, anisotropy, and magnetostriction. Current quantum-mechanical spin-wave theory (Van Kranendonk and Van Vleck, 1958), even though it evades many of these basic problems by using a simplified quantum model, encounters so many mathematical difficulties of quantum nature that it pays scant attention to such other matters as boundary conditions and the shape-dependence of magnetic forces. It will no doubt be a long time before any one theory undertakes to take account simultaneously of all these factors.

The domain and wall concepts have been useful but have sometimes been used with insufficient appreciation of their limitations. On the other hand, the more fundamental methods of micromagnetics have led to solution of only a few problems, and those relate to idealized cases. A reasonable hope is that such methods may be used to establish reliable criteria for the applicability of approximate methods, to provide estimates of the errors in such methods, and to suggest new types of approximation. For this purpose, such concepts as magnetization curling and magnetiza-

tion buckling, which have sprung naturally from the rigorous theory, may be useful. The domain-wall model is properly regarded as the result of imposing certain constants in the equations of micromagnetics; this point of view should be useful in evaluating it and alternative models in various specific applications.

Beyond this, one critical calculation is still needed. The basic postulate of domain theory is that the relevant forces are the exchange, magnetostatic, anisotropy, and magnetostrictive forces. Until rigorous theoretical methods, based on this postulate, have predicted at least one stable domain structure, there can be no assurance that the postulate is, in fact, adequate to account for all of the experimental facts.

4.2.4. Experimental Techniques

One of the most sensitive features of a magnetization loop is the coercive force. A discussion of this quantity will illustrate many aspects of the experimental understanding of structure-sensitive properties. The coercive force of "pure" iron may vary by a factor of over a million depending on structure-sensitive parameters. The object of experimental research in this field is to correlate the value of the coercive force with the chemical and physical structure of the sample. In no instance has this been done in detail.

Fine Particles—As discussed in the previous section, the theory of the magnetization of fine particles is approaching some maturity for simple configurations. In spite of great advances, experiments still deal with particles of diverse shape and size whose properties are not presently amenable to analysis (Luborsky, Mendelsohn and Paine, 1957). Measurements of the rotational hysteresis (W_r) and the rotational hysteresis integral

$$\left(\int_0^\infty (W_R/M_s) d(1/H) \right)$$

indicates, for instance, that the reversal process is not one of uniform rotation (Jacobs and Luborsky, 1957) but, as discussed earlier, the exact nature of the reversal process is unknown.

Thin Films—Both the experimental (Collins and Heavens, 1954; Tiller and Clark, 1958; Lloyd and Smith, 1959; Behrndt and Maddocks, 1959) and theoretical (Néel, 1956) situations regarding domain wall motion in thin films are somewhat primitive. Direct observations of the physical state of the films and the process of wall displacement are badly needed.

Bulk Materials—It is known that strains and impurity aggregates impede the motion of domain walls. Strains have never been

characterized well enough experimentally to deduce their effects. The size and distribution of precipitate particles have rarely been measured. A significant experiment by Dijkstra and Wert (1950) showed that the effect of Fe_3C particles on the coercive force was maximum when the particles had, on the average, a diameter approximately equal to the domain wall thickness.

In addition to the motion of domain walls, there is the problem of nucleation. Nucleation fields very close to the theoretical maximum have been demonstrated in some sections of iron whisker crystals (DeBlois and Bean, 1959). The importance of surface pits has been demonstrated by Bates and Martin (1953, 1956). Williams and Goertz (1952) showed that a notch in a ring of Perminvar reduced the nucleation field by a factor of two. But detailed understanding of wall nucleation at imperfections must await experiments in which the materials and conditions are more precisely controlled.

Conclusion—Techniques are available for reducing impurities to less than one part per million (Pfann, 1957R), for growing crystals with controlled amounts of dislocations or even with no dislocations (Dash, 1959), as well as for producing uniform surfaces (Brenner, 1956). In addition, advances in electron microscopy (Nicholson and Nutting, 1958) and X-ray diffraction and scattering (Beeman, et al., 1957) allow quite exact inference of the physical state of a crystal. These methods have rarely been used in the study of structure-sensitive magnetic properties. In view of the overwhelming technological importance of structure-sensitive properties, further work in this area is certain to have great impact.

5. DYNAMIC PROPERTIES OF FERROMAGNETIC AND FERRIMAGNETIC MATERIALS

5.1. Ferromagnetic Resonance

Any ferromagnetic or ferrimagnetic material will exhibit a magnetic dipole resonance in which the entire spin system precesses as a unit about the direction of an applied field. This phenomenon of ferromagnetic resonance was first observed by Griffiths in 1946, and has been an active field of research since that time. In the usual experiment, a magnetic sample is placed in a microwave resonant cavity between the poles of an electromagnet. The magnetic field is varied so that the spin system passes through resonance with the microwave radiation in the cavity. The position, width and strength of the resonance are the

measured data. The results depend on such parameters as temperature, frequency, sample shape and orientation, and microwave power level. A good discussion of such an experiment has been given by Yager, Galt, Merritt and Wood (1950). A general presentation of the theory of ferromagnetic resonance may be found in a paper by Kittel (1951) although this account is now somewhat out of date. In recent years the situation has been complicated by the discovery that observations can be made on other modes of motion of the spin system than the uniform precession mode. Ferromagnetic resonance experiments yield information about the fundamental dynamics of collective spin systems, about the interaction of the spins with the crystal lattice, and about the imperfections of the spin system. In the discussion below we shall discuss first ferromagnetic insulators such as the ferrites and garnets, and secondly the ferromagnetic metals.

5.1.1. Ferromagnetic Insulators

Our discussion of ferromagnetic resonance in insulating materials will cover a variety of topics that have received attention in recent years and which promise to continue as important lines of research in the future. Resonance measurements upon the ferromagnetic insulators have generally been carried out with one of two ends in view. The object has either been to obtain new information about ferromagnetism of a fundamental nature, or technical information directly useful in design of microwave circuit elements. In the latter case the experimental material has usually been a polycrystalline ceramic. In the search for basic knowledge, however, the use of single crystal materials and relative freedom from imperfections have been absolutely essential. For this reason, we cannot overemphasize the importance to this field of advances in the art of growing single crystals of compounds of the transition metals and rare earth metals.

Of equal importance is research directed toward the development of new magnetic materials. Early resonance work on insulators was limited necessarily to the various ferrites. Although much valuable work was done, tremendous impetus was given to the field by the discovery of the rare earth garnets which possess a very high resistivity and maintain a higher degree of stoichiometry. We may expect similar advances to follow the advent of other new materials if they may be produced in single crystal form and with well-defined chemical composition, stoichiometry, and crystalline perfection. The preparation of ferrimagnetic

fluorides by Knox and Geller (1958) seems to be a promising advance.

We should also remark that there is increasing evidence as noted below that very small amounts of various impurities may be of importance in the resonant response of some of the ferrimagnetic oxides. In ferrites, for example, many properties are very sensitive to the partial interchangeability of atoms on tetrahedral and octahedral sites.

Anisotropy—The frequency at which a ferrimagnetic spin system is resonant depends not only upon the applied field but also upon the internal anisotropy field. It was recognized very early (Van Vleck, 1937; Keffer, 1955) that ordinary dipolar and pseudo-dipolar forces could lead to an anisotropy field. The situation was clarified, however, when a theory of anisotropy based upon the crystalline field splittings of individual ions was proposed by Folen and Rado (1956) and given quantitative theoretical form by Yosida and Tachiki (1957) and by Wolf (1957). This theory has been used to account rather successfully for the temperature-dependent anisotropy of a variety of materials. These theories, however, require the sign of the crystalline field splitting parameter, a , for trivalent iron to be different on tetrahedral and octahedral sites and this is in contradiction to the theory of Watanabe (1957). More work is needed on crystalline field splittings, particularly of trivalent iron on tetrahedral sites. It is possible that consideration of 4th order trigonal or tetrahedral terms in the spin Hamiltonian will help to resolve this difficulty. It must also be noted, as pointed out in section 2.1 on paramagnetism, a purely electrostatic approach to crystalline field splitting may be inadequate.

A special problem is presented by the low temperature anisotropy of yttrium iron garnet which has been found in some samples to be very large and to have an anomalous angular dependence (Dillon, 1958A). More work on different crystals is required to explain this effect. It is possible that in this case, and in others, small amounts of rare earth impurities are important and that more careful control of materials will be essential to progress (Kittel, 1959).

Understanding of the anisotropy field would be helped by precise measurement and comparison of static and resonant anisotropies on carefully selected materials, since these two quantities should be identical at zero degrees Kelvin according to the crystal field model, but are not the same for pseudo-dipolar interactions.

Line Width—The width of the absorption line is one of the

important features of a ferromagnetic resonance experiment. At this time, two sources of line width have been well identified. The first is the valence exchange mechanism proposed by Yager, Galt and Merritt (1955) to account for the line width in certain nickel-iron ferrites. The second mechanism (Clogston, Suhl, Walker and Anderson, 1956; Clogston, 1958) attributes the line width to an inhomogeneous broadening brought about by imperfections in the crystal lattice. The latter effect seems to reside chiefly in the surface of specimens prepared by the usual polishing procedures. This has been vividly demonstrated by LeCraw, Spencer and Porter (1958A), who drastically reduced the line width measured in yttrium iron garnet by careful polishing.

Measured line widths are now so small that it is possible to make preliminary observations on the intrinsic, temperature-dependent line widths predicted by Van Vleck (1950) and Kasuya (1954). Such observations can presently be made only at elevated temperatures where the intrinsic line widths are greatest and inhomogeneous line widths least (LeCraw and Spencer, 1958). It is important that progress be made in polishing and etching techniques so that inhomogeneous widths can be sufficiently reduced to allow observations of the intrinsic line width over a wider temperature range.

A feature of the observations on line width has been the discovery of a strong low temperature peak in manganese ferrite (Seavey and Tannenwald, 1956) and in yttrium iron garnet (Dillon, 1957). The peak has remained a mystery until very recently when Kittel (1959) has offered an explanation based on the presence of rare earth impurities in yttrium iron garnet. Measurements are needed on materials of controlled chemical purity and surface polish over a range of frequencies to test this theory. If the impurities can be established as a source of width, we may expect to see the widths greatly reduced by chemical control of the constituent materials. This could lead to a greatly expanded importance of the field and to radically new technological applications.

A feature of the inhomogeneous line-width broadening discussed above is that the effect should vanish at frequencies sufficiently far below the frequency for ferromagnetic resonances. Since absorption in this frequency range must be ascribed to other mechanisms, such measurements could be helpful in identifying sources of resonance broadening.

Finally, it may be noted that resonance line width in polycrystalline materials is well understood through the work of Schlömann

(1958) who has carefully analyzed the effects of random grain orientation and mutual dipolar interactions of crystallites.

Magnetostatic Modes and Spin Waves—A new feature of ferromagnetic resonance experiments reported in recent years is the observation of modes other than the uniform precession mode. In one class of experiments, these so-called magnetostatic modes are excited by nonuniform driving fields (Solt, White and Mercereau, 1958; Dillon, 1958B). A theoretical analysis of the modes has been made by Walker (1957), and agreement between theory and experiment seems to be satisfactory. More experimental work on these modes will be desirable, particularly in regard to measurement of the line width and saturation of individual modes. Because of mathematical difficulties, the interactions of the magnetostatic modes with each other and the uniform precession mode are not well understood, and further theoretical work along these lines would be justified.

Another interesting observation of magnetostatic modes or spin waves has been made by Jarrett and Waring (1958). In this case the modes are observed in very thin crystals and are apparently excited by virtue of surface effects rather than by nonuniform fields (Kittel, 1958A). This work is new and will certainly lead to new understanding as further experiments are carried out.

Some particularly interesting suggestions have been made by LeCraw and Spencer (1958). By observing the saturation of ferromagnetic resonance, they measure the effective line width to be associated with certain very short wavelength spin waves. This width is observed to be independent of the surface effects noted previously. It is very important to be able to measure in this way an intrinsic property of the spin system. More measurements of this sort, particularly as a function of temperature and frequency and with controlled chemical composition will be very desirable.

Another sort of direct measurement on spin waves becomes possible by the use of high frequency acoustic excitation of a crystal. It has been pointed out by Kittel (1958B) that the interaction of acoustic waves and spin waves will be particularly strong when the waves coincide both in frequency and in wavelength. If the wavelength-frequency relationship of the acoustic waves is known, and if the coincidence condition is determined by observations of the attenuation of acoustic waves as a function of frequency, it enables a calculation to be made of the exchange constant of the spin system. Measurements of this sort should be made.

A related suggestion has been made by LeCraw and Spencer (1958). They are able to measure the effective line width of specific spin waves. If these spin waves become coincident with acoustic waves in frequency and wavelength, a considerable effect on their line width should be observable. An experiment of this sort will also lead to a determination of the exchange constant.

An interesting possibility is raised by the very narrow line widths being currently reported for the uniform precession mode and various magnetostatic modes. With substantial further reductions it may become possible to devise experiments that will directly observe the velocity of propagation of the magnetostatic modes.

High Power Effects—The resonant response of a ferromagnetic body changes radically when the microwave driving power increases above a critical level. Effects of this sort were first noticed by Damon (1953) and Bloembergen and Wang (1954). Two main effects are observed: (1) a premature saturation of the ferromagnetic resonance; and (2) an anomalous absorption of power known as the subsidiary resonance, occurring at a magnetic field lower than that required for resonance. These two effects are now well understood through the theoretical work of Suhl (1957). Additional experimental work by Suhl (1956), LeCraw, Spencer and Porter (1958B), and Spence, LeCraw and Porter (1958) have confirmed predictions of the theory. Additional theoretical work has been done on effect (1) by Schlömann (1958, 1959) that improves the agreement with experiment. Further theoretical work is needed to clarify the nature of the saturated state finally reached by application of very high power and to improve understanding of the threshold level for effect (2). Further experimental work is needed on different materials performed over a range of temperatures and frequencies to further test the theory. Particular attention should be given to conditions that bring into play the acoustic waves mentioned previously.

A further quantity that can be measured in ferromagnetic resonance by the application of large signal powers is the rate of relaxation of the component of magnetization directed along the applied magnetic field. This measurement determines the rate at which excitation passes out of the total spin system and is important in disentangling interactions of the spin system with the crystal lattice from internal interactions of the spin system. New experiments have been reported recently by Farrar (1958) but more work is needed.

5.1.2. Ferromagnetic Metals

We turn now to a consideration of ferromagnetic resonance phenomena in metals and alloys. We will cover here again many of the topics discussed in connection with the ferromagnetic insulators. In general, although much important new understanding has been gained in recent years, we shall see that understanding of resonance in the metals is not as far advanced as in the case of the insulators. This is due in part to the more complicated nature of some of the phenomena, in part to the difficulty of obtaining sufficiently perfect single crystal specimens, and in considerable measure to the fact that the electronic structure of the transition metals is still very obscure.

General Remarks—Although ferromagnetic resonance measurements can be made on polycrystalline metallic specimens, the corrections required to reduce the data often make the results of dubious value. Experiments made on single crystals are therefore important in determining g -values and anisotropy constants. Similarly, measurements of line width are only significant when performed on single crystal specimens (except in some special cases noted below). The principal measurements of this sort have been made by Kip and Arnold (1949) and Barlow and Standley (1956) on silicon-iron crystals, and by Reich (1956) on nickel crystals. There is, comparatively speaking, a great paucity of available data. Progress in the understanding of ferromagnetic resonance in the metals is going to depend upon the carrying out of many more experiments on single crystals of different ferromagnetic metals and alloys over wider ranges of temperature and frequency. Of particular importance will be experiments made on very perfect single crystals such as reported by Rodbell (1959).

Anisotropy and g -value—The anisotropy constants determined from the single crystal measurements referred to above have been found to agree fairly well with static measurements made on similar crystals. The current theory of anisotropy and g -value in the ferromagnetic metals was first given by Brooks (1940) and extended by Fletcher (1954) based, to a first approximation, on the broadening of the atomic levels into bands. This contrasts with the theory of Yosida (1957) and Wolf (1957) for the non-metals in that the origin of the anisotropy is here ascribed to the splitting of d levels in the crystalline electric field. The theory is unsatisfactory in several respects and requires serious revision. It predicts a value for the first anisotropy constant, K_1 , of nickel about 50 times larger than the experimental value, and a g -value of 2.14 which may be compared to the experimental value of 2.22 from the

work of Reich (1956). It is clear that much more theoretical and experimental work is needed. The results will be important in clarifying the electronic state of the transition metals. It will be desirable, just as for the nonmetals, to make a close comparison of the static and resonant anisotropies on identical crystals to see if any discrepancies exist. It may be noted in this connection that Reich finds a value of the second anisotropy constant K_2 opposite in sign to that found from static measurements. Some very interesting experiments made with small particles of cobalt precipitated in copper that bear on this question have been carried out by Bean, Livingston and Rodbell (1958).

It would appear particularly important that measurements be made on the magnetic rare earth metals, especially gadolinium. The only measurements so far reported (Kip, 1953) were not made on single crystals. The physical mechanisms operating in the transition metals and in the rare earth metals may be sufficiently different so that the comparison of their properties will be extremely helpful.

Line Width—The line widths observed for ferromagnetic resonance in the metals have been generally of the order of hundreds of oersteds, very much wider than the (<1 oersted) lines being currently obtained with the nonmetals. Many suggestions have been offered to account for the line width, but they generally lead to much too small an effect. The most important mechanism was proposed by Kittel and Mitchell (1956) and presented in detail by Mitchell (1957). This theory accounts for the line width by an exchange interaction between s and d electrons and rapid relaxation of the conduction electron spins. Although the calculated line width is an order of magnitude greater than that predicted from previous theories, it is still an order of magnitude less than the line width observed in nickel by Reich (1956).

Before discussing this situation further, we must mention a totally different line broadening mechanism. Rado and Weertman (1954) carried out ferromagnetic resonance measurements on carefully prepared, low anisotropy alloys of nickel and iron. The line widths they obtained (judged in their case by the susceptibility at resonance) were very much less than in any previous experiment. Furthermore, they were able to account for the width and shape by a mechanism that converted spin energy into ohmic loss of conduction electrons by means of the exchange torques developed in the skin depth layers of the metal. It appears that all other mechanisms were reduced to unimportant proportions by the choice of a low anisotropy alloy. In addition, Rodbell (1958)

has found in his experiments on iron single crystal whiskers that the line width can be accounted for almost entirely by the exchange-conduction electron mechanism. These two observations suggest strongly that much of the observed line width in the usual experiments may be inhomogeneous broadening brought about by crystal imperfections acting through anisotropy of magnetostrictive fields.

It would appear, therefore, that future experiments should concentrate on the low anisotropy alloys or on very perfect metal whisker crystals. It will be desirable that line widths be explored in different materials over broad ranges of temperature and frequency. The exchange-conduction electron mechanism for example is in difficulty at lower temperatures (Weertman and Rado, 1958) and it will be desirable to improve the extended theory of Rado (1958) which includes the anomalous skin effect. Furthermore, it should be possible to find circumstances under which the exchange-conduction electron effect is negligible in order to test the *s-d* exchange mechanism of Kittel and Mitchell.

Thin Films—In recent years, due to the need for fast switching devices, there has been a growing interest in the ferromagnetic resonance behavior of metal films. Much of the work has been done by Tannenwald and Seavey (1957, 1958) who have used thin films of Permalloy evaporated on a glass substrate. The line width observed in these experiments can be completely accounted for by the exchange-conduction electron effect described above. This result presumably arises from the use of the low anisotropy Permalloy film. A particularly interesting observation is that the lines are not broadened by obvious, large-scale inhomogeneities in the films indicating the effectiveness of some long-range (presumably dipolar) coupling force. As these experiments have been carried out by Tannenwald and Seavey, a measurement has also been made of the magnetization, *M*, which permits the determination of the exchange integral *J*. Further experiments of this sort, particularly on ultra-thin films, will contribute to the general understanding of ferromagnetic resonance in metals.

A feature of the resonance experiments in thin films is the appearance of a series of subsidiary resonances on the low-field side of the main resonance. These resonances arise from the same mechanism of surface pinning proposed by Kittel (1958) to explain the results of Jarrett and Waring (1958) in thin crystals of ferromagnetic insulators. The experiments are valuable in making available to observation specific spin waves, and also provide

important information about the interactions of spins with the surface of the film.

An interesting application of ferromagnetic resonance has been made by Seavy and Tannenwald (1957, 1958) to the determination of the magnetization of thin films. The measurement is carried out by determining the resonance frequency with the applied field parallel and perpendicular to the film and makes possible the study of magnetization as a function of film thickness in extremely thin films. The agreement between theory and experiment is poor and much more work is required.

Microwave Hall Effect—In the course of ferromagnetic resonance experiments in thin films, a new effect has been observed and reported by Tannenwald and Seavy (1958) and by Eagan and Juretschke (1958). In these experiments, a modulated d.c. signal appears on two wires soldered to the film when the film is set into ferromagnetic resonance by a large, pulsed microwave signal. The effect is apparently due to the interaction of the radiofrequency currents in the films with the out-of-phase components of magnetization. It is not yet clear whether study of this effect will lead to new understanding of a fundamental nature, but further experiments carried out on a variety of materials under various experimental conditions are clearly desirable.

5.2. A. C. Losses

The measurements of losses associated with a finite rate of magnetization are made in many ways (e.g., Bean and Becker, 1959). In general, they fall into two classes. The first deals with variations in the magnetization so small that the magnetization can be described by a permeability; in particular, the imaginary part of the permeability is measured. The second class comprises measurements on non-linear changes of magnetization with field. Here the measurement methods range from calorimetric measurements to measurements of the rate of reversal caused by sudden application of a magnetic field. One of the most illuminating methods of experiment in this area has not received much attention. It is the measurement of the torque on a magnetic sample produced by a uniform, rotating magnetic field (Gilbert and Kelly, 1955; Kelly, 1956). The feature of this measurement, in contrast to the usual type, is that the state of the magnetization is well specified, i.e., it is undergoing substantially uniform rotation.

The understanding of the losses that accompany large-scale changes of the magnetization is most complete for bulk metals in

which the main loss mechanism is the direct generation of eddy currents. As mentioned in the previous section, experiment and theory for single domain walls are in good agreement. Formal extension of this theory to arrays of domains (Polivanov, 1952; Pry and Bean, 1958) meets no difficulties but experimentally (Hu and Weiner, 1959; Neurath, 1959; Carr, 1959) the problem exists of knowing how many domain walls are moving in a given exciting field.

For thin metal films, other ill-defined losses become apparent (e.g., Goodenough, 1957). These mechanisms form the major source of loss in thin films (Smith, 1958). For large-scale changes of magnetization the losses in thin metals, ferrites and garnets have been found by Gyorgy (1958) to be surprisingly similar to one another. It seems likely that structure-sensitive properties are playing a role here, and that theories based on a perfect lattice have no relation to existing experiments. One area in which we have some detailed understanding is that of electron transfer in nonstoichiometric ferrites (Galt, 1954, 1955; Clogston, 1955). Experiments on garnets should be very revealing in this connection since this is one case in which electron transfer does not take place.

The topic of losses under large-scale excitation is of intense technological importance for fields as diverse as power transformers and computer elements. Fundamental progress may be expected to be made by experiments on pure materials of high perfection and by extension of the theories of spin-lattice relaxation to include the defects existing in real crystals.

5.3. Magnetic After Effects

A magnetic material does not respond instantaneously to an applied field. The basic experiment may be performed in one of two ways. Either the field can be suddenly changed and the subsequent course of the magnetization traced as a function of time, or the material can be subjected to a small alternating field. In the latter case one can measure the real and imaginary parts of the permeability as a function of frequency. The time effects arise from a great diversity of causes and may have complex interactions with one another. The experimental and theoretical picture is clear in only a few cases. The principal reason for this lack of clarity arises from the fact that these are structure-sensitive phenomena, and experiments have, in most cases, been carried out on materials whose imperfections have not been characterized.

Eddy Currents—Voltages induced by changing magnetization cause currents to flow in a metal. These currents cause two related

effects. First, they dissipate energy by Joule heating, and secondly they produce magnetic fields in opposition to an applied field. The latter effect causes the field, at some interior point of the metal, to approach more or less slowly the value of the applied field in a pulse experiment. With alternating fields, the magnitude and phase of the field varies through the specimen. In principle, if one knows the response of the local magnetization to magnetic fields and the resistivity of the material, one can, by use of Maxwell's equations, derive the magnetic behavior as a function of time for arbitrary applied fields. There are, however, formidable mathematical difficulties. Solutions were first obtained for simple geometric shapes and with the assumption of a uniform permeability (Bozorth, 1951). The use of a uniform and constant permeability is dangerous because of the essential heterogeneity of the magnetization process (see the previous section on Structure Sensitive Properties). In some cases, where the domain size is small compared to sample dimensions and where small fields are applied, the approximation is quite good (Wweden-sky, 1921). Often, however, the macroscopic permeability is not uniform (Peterson and Wrathall, 1936; Epelboin, 1951). On the other hand, solutions have been given for simple domain geometries that, at least to first order, take into account the heterogeneity of magnetization (Williams, Shockley and Kittel, 1950; Néel, 1952). These solutions give, for single walls, reasonable agreement with experiment and have recently been used as a tool in determining dynamic wall configurations (Becker, 1959). A deduction from these experiments (Rodbell and Bean, 1956) is that wall motion is essentially dissipative, and therefore models (Kersten, 1943) that postulate a conservative potential function of wall energy versus displacement are essentially unrealistic.

Thermal Fluctuations of Magnetization—Since the local process of magnetization is often characterized by irreversibility, one might expect thermal fluctuations to cause the magnetization to surmount potential barriers in a way dependent on the barriers, field, temperature and time. The clearest physical picture of one form of this process is given by consideration of a single domain particle at elevated temperatures (Néel, 1949). The barrier is the same one that gives rise to the coercive force (see sec. 4.2). Thermal fluctuations of strain or shape cause the magnetization to precess over the barrier (Brown, 1959). While this effect has been observed (Weil, 1954; Bean and Livingston, 1959), the experimental particles have never had sufficient uniformity to critically test this theory. Measurements on Alnico type permanent

magnet materials (Street and Woolley, 1949, 1950, 1956) have usually been interpreted on the basis of thermally-activated wall motion but may be more closely related to the effect described above.

In bulk material, thermal fluctuations will cause domain walls to surmount local barriers. In a rather untransparent theory (Néel, 1950, 1951), the time change of magnetization is related to the hysteresis loop as well as time, temperature and field. The most significant aspect of experiment (Brissonneau, 1958) and theory is that there is a wide spread of activation energies, and hence the dependence on time and temperature is not strong. Experiments on single crystals with defined inhomogeneities would be useful.

Thermal Motion of Barriers—Closely allied to the above considerations is that the barriers themselves may diffuse or migrate. One case that is becoming increasingly well understood is that due to interstitial motion in b.c.c. metals. Specifically carbon in iron (reviewed by Rathenau, 1959; Néel, 1959) has been shown to have a specific magnetic interaction with the magnetization of the iron lattice. Thus, if given time, the carbon atoms within a 180° domain will take up positions that minimize their free energy. The ease of subsequent motion of the wall and the losses occasioned by the motion are governed by the time scale of the motion relative to the jump time of the carbon atoms. The activation energy is in accord with direct measurements, and, indeed, this effect gives a method for measuring diffusion constants (Bosman, Brommer, Van Daal and Rathenau, 1957). In addition, from the distance that the wall must be moved quickly before it leaves the potential well it created, the thickness of the wall may be inferred (Rathenau, 1959). This thickness is in good accord with computed values. Essentially similar effects to those of interstitial rearrangements are found in atomic and electronic (Clogston, 1955) rearrangements in ferrites. A related effect occurs in substitutional alloys where a directional short-range order of the constituents may be produced by interaction with the magnetization (Néel, 1953; Taniguchi and Yamamoto, 1954). This interaction is thought to be one source of magnetic annealing (reviewed by Graham, 1959), but other factors may be important, at least in nickel-iron alloys (Heidenreich, Nesbitt and Burbank, 1959).

Other Causes—There are many other possible sources of magnetic aftereffects. They cover the same range as the causes of mechanical aftereffects (Zener, 1948). Possibilities include: induced dislocation motion, grain boundary creep under magneto-

strictive stresses, and thermal effects due to non-uniform flux change or magnetostriction.

Conclusion—While some of the main elements of magnetic after-effects are becoming clear, extensive work on single crystals of controlled perfection will be necessary to test critically various concepts and expand our understanding.

6. DIAMAGNETISM AND SUPERCONDUCTIVITY

Diamagnetism is an effect which has been used only as a tool for the study of materials—notably metals, through the de Haas-van Alphen effect. Usually, it is so small as to be of little importance, but occasionally, as in Bi, graphite, or some large aromatic molecules, it has an appreciable magnitude. Diamagnetic or cyclotron resonance has also been of great value in the physics of semiconductors and increasingly so for metals. Technical uses have not as yet appeared.

Superconductivity, on the other hand, is a striking effect whose technological usefulness is limited only by the fact that, as yet, it has been found to occur only below liquid hydrogen temperatures. It is classed as a magnetic effect because the most fundamental property is the Meissner effect: a portion of a metal in the superconducting state will allow no magnetic field, B , except in surface layers $\approx 10^2 - 10^3$ Å thick. Infinite conductivity is also observed; in simple situations these two are equivalent.

Some obvious direct applications of infinite conductivity, useful right now in experimental work, and perhaps more so in the future when liquid helium technology has improved still further, are: (1) a superconducting loop as a permanent magnet; (2) superconducting cavity (both may be of use in maser technology); (3) galvanometer coil; (4) transmission of information, where it is not economically ridiculous compared to present repeatered lines (Galt, private communication). Special properties lead to the obvious: (5) magnetic shielding, though one must realize that only the a.c. magnetic field is shielded; (6) heat switches at low temperatures; (7) the nonlinearity at the transition field leads to bolometer and amplifier possibilities. Most promising of all at present is the computer memory possibility using flux-trapping schemes, where speed and ease of construction are favorable (Garwin, 1957).

Superconductivity should also not be ignored as a tool in the physics and chemistry of metals. Particularly, like any such transition, it is a sensitive device for studying metallurgical phase diagrams; for example, the recent revision of the solubilities of

the iron group in Ti and Zr (Matthias, 1955; Matthias and Corenzwit, 1955; Matthias, private communication). No doubt the theoretical understanding we are near reaching will point out many other uses in metal physics.

Many good reviews on the "classical" or pre-B.C.S. (Bardeen, Cooper, and Schrieffer, 1957)* aspects of superconductivity exist, notably the books by Shoenberg (1952R) and F. London (1950R). A review primarily of superconducting materials has been given by Matthias (1957R).

Bardeen, Cooper, and Schrieffer (1957) recently proposed the first promising microscopic theory of superconductivity, and, both theoretically and experimentally, the field is dominated by the attempt either to exploit and extend this theory or to find a definitive proof or disproof.

This theory begins with Cooper's observation that attractive forces between electrons cause the electron Fermi gas, which we imagine the normal metal to contain, to be unstable at low enough temperatures. Coupling to the lattice can cause the attractive force: one electron polarizes the lattice near itself in such a way that a second electron also "likes" that vicinity. The instability occurs because a pair of electrons at the Fermi surface can form a bound state. The wave function of such a bound state is $\psi(r_1 - r_2) = \int f(k) e^{ikr_1} e^{-ikr_2} e^{-ikr_2}$, so it is a linear combination of $+k, -k$ pairs. Because there are a large number of possible k -values for such pairs, any attractive force, however weak, can cause the bound pairs to form.

The B.C.S. ground state forms the whole surface of the electronic Fermi sea out of such bound pairs. One may think of the energy gap in two ways: first, that there must be such a gap in the density of states so that it is not energetically favorable for any more bound pairs to form, or second, that the gap is the binding energy of a pair, and that individual electrons can only run free by breaking up a pair and losing this energy. The essential achievement of B.C.S. was to demonstrate how such a state could be formed, and how it leads to an energy gap.

Most of the properties of superconductors can be correlated with the existence of a gap in the energy spectrum of the sort predicted by the B.C.S. theory. It must be of this particular sort; a semiconductor, for instance, has a gap, but one associated with the existence of the lattice of ions, so that the electrons cannot

*Hereafter, the work of Bardeen, Cooper, and Schrieffer (1957) will be referred to as B.C.S.

move freely with respect to the ion lattice. There are also certain coherence properties, which make the acoustic absorption, say, or the spin susceptibility drop to zero quite differently from nuclear relaxation as the gap opens up.

The typical superconducting properties follow from this state in much the same way first envisaged by London. (Difficulties pointed out by Schafroth (1958) in this derivation have been finally and decisively settled (Anderson, 1958A, 1958B; Pines and Schrieffer, 1958; Rickayzen, 1958).) London's idea is equivalent to supposing that the electron gas in the presence of fields accelerates as it would if it were a rigid whole and no scattering took place. In particular, it would undergo the Larmor precession in the presence of a magnetic field. This precession current causes, because of the macroscopic size of the sample as compared to the atoms in which it usually occurs, enormous magnetic fields which cancel out the applied fields leading to $B \equiv 0$. In general, such a gas behaves just like a superconductor does.

With extensions to carry the theory to the same stage as the modern "highbrow" theories of the free electron gas, (Anderson, 1958B; Pines and Schrieffer, 1958; Rickayzen, 1958) the situation is hopeful from a qualitative point of view. That is, we suspect that we understand superconductivity as a state of matter, like the crystalline, say, or antiferromagnetic states, without having a very clear idea of the forces which cause it, without rigorous proofs that it really is the stable state, and without any complete criterion for its occurrence in one material rather than another. But we have no such information on these other states of matter either.

Alternative possibilities have been proposed by Fisher (private communication) and others (Heine and Pippard, 1958; Schrieffer, private communication) mostly because of the Knight shift difficulty soon to be discussed. These theories seem most unlikely to apply to the majority of superconductors, because the assumed interactions are unlikely, and various experimental facts are not explained.

Experimentally, the general picture is consistent with B.C.S. and, particularly, with the energy gap for one-particle excitations it contains. The most serious result which clearly has seemed to contradict it is the observation of a finite Knight shift in Hg (Reif, 1957) and Sn (Knight, private communication) at 0°K. The B.C.S. superconductor can have no spin susceptibility at 0°K, because the electron pairs are in singlet states. Therefore, the application of a magnetic field will not lead to a polarization of

the electron spins that would be reflected in a shift of the nuclear resonance. On the other hand, the experimental situation is not that of a pure bulk superconductor; the particles are fine and the measurements are within ~ 100 Å of a surface. Suhl (1957) has pointed out that the Knight shift is a function of position in a bulk metal because the susceptibility is nonlocal—i.e., the shift at position r depends on the field at all points r' nearby—while Ferrell (1959) has emphasized that this, together with spin-dependent surface or volume scattering caused by spin-orbit forces, can lead to a Knight shift. The latter type of effect certainly occurs strongly in the metals so far tested. Surface scattering depends on atomic number strongly and more experiments, especially on lighter metals and different samples, are very much indicated. The technique developed by Knight for this experiment also is of great interest for the study of finely divided materials in general.

A rather serious puzzle is also posed by the thermal conductivity as far as electron-phonon scattering is concerned; (Bardeen, Rickayzen and Tewordt, 1958) but present acoustic absorption measurements are a strong argument for B.C.S. (Bardeen, Cooper and Schrieffer, 1957).

Deviations of the specific heat curve from the purely B.C.S. form (Decker, Mapother, and Shaw, 1958) are to be expected in real metals where the energy gap is probably quite anisotropic, and thus such measurements do not disprove the theory, which in any real metal would lead to considerable anisotropy. Recently acoustic measurements have confirmed this anisotropy (Morse, Olsen and Gavenda, 1959). Correlation of these measurements with the scattering theory and with band shapes are of importance (Phillips, 1958). Not all deviations of specific heat curves are easily explained this way; also, some of them, as well as the existence of exciton states, seem to correlate with low Debye temperature (Anderson, to be published).

Measurements of nuclear relaxation (Hebel and Slichter, 1957; to be published) in Al showing the pile-up of states near the gap, and direct infrared and microwave measurements of the gap (Biondi, Forrester, Garfunkel and Satterthwaite, 1958; Ginsburg, Richards and Tinkham, private communication) have remarkably confirmed B.C.S. Recent observations of structure may well indicate excitons (Tinkham, 1959). Such optical measurements must necessarily be carried out on very bad specimens, and it is to be hoped that some more satisfactory spectroscopy can be worked out in the future. The B.C.S. theory leads to a Pippard

nonlocal electrodynamics, and this agrees well enough with experiment also, as well as with the general arguments of Ferrell et al. (1958), (Tinkham and Ferrell, 1959) based on certain sum rules.

Perhaps the most convincing arguments in support of B.C.S. come from considerations of time-reversal and scattering. Magnetic scattering, which breaks up B.C.S. pairs, in some cases destroys superconductivity where a comparable amount of elastic ordinary scattering does not, as can be explained by a modified B.C.S. theory (Phillips, 1958). On the other hand, some magnetic additions do not have this effect, and even cases of ferromagnetism, as well as antiferromagnetism accompanying superconductivity in the same specimen, have appeared (Matthias, Suhl and Corenzwit, 1958). Both experimental and theoretical work in this area are of the greatest importance.

Finally, transition temperatures and energy gaps have been calculated from first principles by Pines (1958) and Morel (1958). While these calculations are as accurate as most melting temperature or Curie point calculations, they leave a great deal to be desired, and in particular do not explain the failure to observe superconductivity above 18°K. Empirical rules about the criteria for superconductivity have enabled Matthias (1957R) to discover hundreds of new superconducting compounds and alloys; we can expect that many more exist. Matthias' rules also give rough (but often surprisingly close) estimates of T_c for new materials, and except for the recent discoveries in magnetic materials, seem to provide a quasi-empirical but complete answer to the question of occurrence of superconductivity. Theoretical attempts to explain these rules have as yet been rather unsatisfactory. Physical measurements on these new materials have only begun to be made and may be of the utmost importance.

A group of subjects which have had less emphasis in the recent period are associated with the "intermediate state": trapped flux, boundary motion, supercooling, etc. Theory has made almost no progress in this technically important field. The vital entity here is the Bloch-type wall between normal and superconducting regions; its wall energy and wall motion dynamics are as yet not understood. The phenomenon of flux trapping ("hard superconductivity"), observed so widely in impure samples, needs further study, as also the dynamics of the magnetic superconducting substances. One may expect critical tests of the theory of B.C.S. here in the future.

In summary, measurements of bulk properties of pure super-

conductors in general agree with B.C.S., and seem a less promising field for the future than the study of controlled impurity samples, small particles, etc. The magnetic superconductors are also a wide-open field.

The general field of research through controlled materials, here as in all of solid state physics, should be emphasized. The few examples yet available (the Rutgers work on chemical impurities in Sn and In; the Bell Laboratory work on magnetic additions in La and Ti) have turned up exciting new effects in each case, and more physical measurements remain to be done even on these examples. As yet even less work has been done on the effect of physical defects in materials.

Finally, we can expect an increasing attempt to correlate intermediate state phenomena with theory.

Note added in proof: Since submission of this work for publication, much has happened in the field of superconductivity—particularly on the experimental side. Noteworthy is the important new work on alloys and intermetallic compounds with extremely high critical fields, measurement of the quantized flux by Fairbank and others and the more recent results on the Knight and isotope shifts which raise some questions on the B.C.S. theory. It would be impossible to bring this material up to date. The reader is referred to the proceedings of various conferences on this subject which were being held as this goes to press.

7. BIBLIOGRAPHY AND REFERENCES

7.1. Bibliography: Books and Review Articles

General

- Becker, R. and Doring, W., 1939, *Ferromagnetism* (Berlin, J. Springer).
Bitter, F., 1937, *Introduction to Ferromagnetism* (N.Y., McGraw-Hill).
Bozorth, R. M., 1951, *Ferromagnetism* (N.Y., D. Van Nostrand).
Goldman, J. E., 1957, *Science of Engineering Materials*, Chapt. 12, 13 (N.Y., John Wiley).

Paramagnetic Resonance

- Bleaney, B. and Stevens, K. W. H., 1953, Reports on Progress in Physics, The Physical Society, London, Vol. 16, p. 108.
Bowers, K. O. and Owen, J., 1955, Reports on Progress in Physics, The Physical Society, London, Vol. 18, p. 304.

Domains and the Magnetization Curve

- Arrott, A. and Goldman, J. E., 1959R, *Fundamentals of Ferromagnetism*, Electrical Manufacturing, March, 1959.

Kittel, C., 1949, *Rev. Mod. Phys.* **21**, 541.

Kittel, C. and Galt, J. K., 1956, *Solid State Physics*: Vol. 3 (New York, Academic Press), pp. 437-564.

Stewart, K. H., 1954, *Ferromagnetic Domains* (Cambridge, University Press).

Stoner, E. C., 1950, *Repts. Progr. Phys.* **13**, 83.

Wohlfarth, E. P., 1959, *Adv. in Phys.* **8**, 30. (Review of Hard Magnetic Materials).

Antiferromagnetism

Nagamiya, T., Yosida, K., and Kubo, R., 1955, *Advances in Physics* **4**, 1.

Masers

Weber, J. A., 1959, *Rev. Mod. Phys.* **31**, p. 681.

Anisotropy

Shenker, H., 1959, in *Methods of Experimental Physics*, K. Lark-Horovitz and V. A. Johnson, eds. (New York, Academic Press), **6B**, pp. 243-247.

Techniques of Crystal Growth

Pfann, W. G., 1957, in *Solid State Physics*, F. Seitz and D. Turnbull, eds., (New York, Academic Press), Vol. **4**, pp. 423-521.

Crystal Field Theory

Orgel, L. E. and Griffith, J. S., 1957R, *Quart. Revs. (London)* **11**, 381.

Ballhausen, C. J., 1954, *Kgl. Dansk. Vid. Selsk. Math. Phys.* **29**, No. 4.

Smit, J. and Wijn, H. P. J., 1959, *Ferrites* (New York, John Wiley).

Microwave Properties of Solids

Lax, B., 1959, *The Microwave Journal* **2**, 9.

Bagguley and Owen, 1957, *Repts. Prog. Phys.* **20**, 204.

Superconductivity

Shoenberg, David, 1952, *Superconductivity*, Cambridge, University Press.

London, Fritz, 1950, *Superfluids*, Vol. 1, Macroscopic Theory of Superconductivity (New York, John Wiley and Sons, Inc.).

Matthias, B. T., 1957, *Progress in Low Temperature Physics*, Vol. II.

7.2. References

Abragam and Proctor, 1958, *Compt. rend.*, **246**, 2253.

Abragam, McCausland, and Robinson, 1959, *Phys. Rev. Letters*, **2**, 449.

Aharoni, A., and Shtrikman, S., 1958, *Phys. Rev.*, **109**, 1522.

Akulov, N., 1931, *Z. Physik* **69**, 78.

Anderson, P. W., 1950, *Phys. Rev.*, **79**, 350.

———, 1957, *J. Appl. Phys.*, **28**, 1049.

———, 1958A, *Phys. Rev.*, **110**, 827.

———, 1958B, *Phys. Rev.*, **112**, 1900.

———, 1958, *Phys. Rev.*, **109**, 1492.

- , 1959A, *Phys. Rev.*, **114**, 1002.
———, 1959B, *Phys. Rev.*, **115**, 2.
———, *J. Phys. Chem. Solids* (to be published)
———, and Hasegawa, H., 1955, *Phys. Rev.*, **100**, 675.
Bacon, G. and Street, R., 1958, *Phys. Soc.*, **72**, 470.
Baker, Winslow and Yager, 1955, *J. Am. Chem. Soc.*, **77**, 4751.
Bardeen, Cooper and Schrieffer, 1957, *Phys. Rev.*, **108**, 1175.
Bardeen, Rickayzen and Tewordt, 1958, (to be published), given at International Conf. on Metals at Low Temperatures, Geneva, N.Y.
Barlow, G. S., Standley, K. J., 1956, *Proc. Phys. Soc. B*, **69**, 1052.
Basov, N. G., and Prokhorov, A. M., 1955, *JETP (USSR)* **23**, 249.
Bates, L. F. and Martin, D. H., 1956, *Proc. Phys. Soc. (London)* **B69**, 145.
——— and ———, 1953, *Proc. Phys. Soc. (London)* **A66**, 132.
Batterman, B. W., 1959, *Phys. Rev. Letters*, **2**, 47.
Bean, C. P. and Becker, J. J., 1959, *Methods of Experimental Physics*, K. Lark-Horowitz and V. A. Johnson, eds. (N.Y., Academic Press), **6B**, Chap. 8.3.
——— and Livingston, J. D., 1959, *J. Appl. Phys.*, **30s**, 120s.
Bean, Livingston, and Rodbell, 1959, *J. Phys. Rad.*, **20**, 298.
Becker, J. J., 1959, *J. Appl. Phys.*, **30**, 387.
Beeman, Kaesberg, Anderegg and Webb, 1957, *Handbuch der Physik*, S. Flugge, ed. (Berlin, Springer-Verlag), **33**, 321.
Behrendt, Legvold, and Spedding, 1958, *Phys. Rev.*, **109**, 1544.
Benedek, G. and Kushida, T., 1960, *Phys. Rev.*, **118**, 46.
Bethe, H., 1933, *Handbuch der Physik*, Vol. XXIV-2 (Berlin, Springer-Verlag).
Behrndt, K. H., and Maddocks, F. S., 1959, *J. Appl. Phys.*, **30s**, 276s.
Bleaney, B., and Bowers, K. D., 1952, *Proc. Royal Soc.*, **A214**, 451.
Bloch, F., 1932, *Z. Physik*, **74**, 295.
Bloembergen, N., 1956, *Phys. Rev.*, **104**, 324.
Bloembergen, N., and Rowland, T. J., 1953, *Acta Met.*, **1**, 731.
——— and ———, 1955, *Phys. Rev.*, **97**, 1679.
———, Shapiro, S., Pershan, P. S., and Artman, J. O., 1959, *Phys. Rev.*, **114**, 445.
———, and Wang, S., 1954, *Phys. Rev.*, **93**, 72.
Bogolyubov, Shirkov, and Tolmachev, 1958, *New Method in the Theory of Superconductivity*, Academy of Sci., U.S.S.R., Moscow.
Bosman, A. J., Brommer, P. E., van Daal, H. J., and Rathenau, G. W., 1957, *Physika*, **23**, 989.
Bowers and Mims, to be published.
Brissonneau, J., 1958, *Phys. Rad.*, **19**, 490.
Brooks, H., 1940, *Phys. Rev.*, **58**, 909.
Brown, W. F., Jr., 1940, *Phys. Rev.*, **58**, 736.
———, 1941, *Phys. Rev.*, **60**, 139.
———, 1957, *Phys. Rev.*, **105**, 1479.
———, 1959, *J. Appl. Phys.*, **30**, 62s.
Carr, W. J., Jr., 1959, *J. Appl. Phys.*, **30s**, 90s.
Chester, Wagner, and Castle, 1958, *Phys. Rev.*, **110**, 281.
Clogston, A. M., 1955, *Bell Systems Tech. J.*, **34**, 739.
Clogston, A. M., 1955, *ibid.* **34**, 759.
———, 1958, *J. Appl. Phys.* **29**, 334.
———, Suhl, H., Walker, L. R., and Anderson, P. W., 1956, *J. Phys. Chem. Solids*, **1**, 129.

- Cochhardt, A. W., 1959, in *Magnetic Properties of Metals and Alloys* (Cleveland, American Soc. for Metals).
- Collins, L. E., and Heavens, O. S., 1954, *Phil. Mag.*, **45**, 283.
- Collins, Kyhl, and Strandberg, 1959, *Phys. Rev. Letters*, **2**, 88.
- Commoner, Heise, and Townsend, 1956, *Natl. Acad. Sci., U.S.*, **42**, 710.
- Commoner, Townsend, and Pake, 1954, *Nature*, **174**, 689.
- Conbrisson, J., Honig, A., and Townes, C. H., 1956, *Compt. rend.*, **242**, 2451.
- Crowe, J. W., 1957, *I.B.M. Jour. Res.*, **1**, 294.
- Danan, H., 1959, *J. de Phys. et Rad.*, **20**, 296.
- Dash, W. C., 1959, *J. Appl. Phys.*, **30**, 459.
- DeBlois, R. W., and Bean, C. P., 1959, *J. Appl. Phys.*, **30s**, 225a.
- Decker, Mapother, and Shaw, 1958, *Phys. Rev.*, **112**, 1888.
- Dicke, R. H., Private Communication.
- Dijkstra, L. J., and Wert, C., 1950, *Phys. Rev.*, **79**, 979.
- Dillon, J. F., Jr., 1957, *Phys. Rev.*, **105**, 759.
- , 1958, *J. Appl. Phys.*, **29**, 539-541.
- , 1958A, *Phys. Rev.*, **111**, 1476.
- , 1958B, *Phys. Rev.*, **112**, 59.
- , and Neilsen, J. W., 1959, *Phys. Rev. Letters*, **3**, 30.
- Dransfeld and Bömmel, *Phys. Rev. Letters*, 1958, **1**, 234; **2**, 298, 1959.
- Dunitz, J. D., and Orgel, L. E., 1957, *J. Phys. Chem. Solids*, **8**, 20.
- , and ———, 1957, *J. Phys. Chem. Solids*, **8**, 318.
- Dyson, F. J., 1955, *Phys. Rev.*, **98**, 349.
- Eagan, W. G., Juretschke, H. J., 1958, *Bull. Amer. Phys. Soc. Ser. II*, **3**, 194.
- Elmore, W. C., 1938, *Phys. Rev.*, **53**, 757.
- Epelboin, I., 1951, *Rev. Gen de L'Electricite*, **60**, 78.
- Farrar, R. T., 1958, *J. Appl. Phys.*, **29**, 425.
- Feher, G., 1956, *Phys. Rev.*, **103**, 834.
- , 1957, *Phys. Rev.*, **105**, 1122.
- , 1959, *Phys. Rev. Letters*, **3**, 135.
- Feher, Gordon, Buehler, Gere, and Thurmond, 1958, *Phys. Rev.*, **109**, 221.
- Feher, G., and Kip, A. F., 1955, *Phys. Rev.*, **98**, 337.
- Feher, G. and Scovil, H. E. D., 1957, *Phys. Rev.*, **105**, 760.
- Ferrell, R. A., 1959, *Phys. Rev. Letters*, **3**, 262.
- , and Glover, R. E., III, 1958, *Phys. Rev.*, **109**, 1398.
- Fletcher, R. C., 1954, *Proc. Phys. Soc. A*, **67**, 505.
- Fletcher, Yager, Pearson, Holden, and Merritt, 1954, *Phys. Rev.*, **94**, 1392.
- Folen, V. J., and Rado, G. T., 1956, *Bull. Am. Phys. Soc., Ser. II*, **1**, 132.
- , and ———, 1957, *Bull. Am. Phys. Soc., Ser. II*, **2**, 263.
- Frei, Shtrikman, and Treves, 1957, *Phys. Rev.*, **106**, 446.
- Galt, J. K., 1954, *Bell System Tech. J.*, **33**, 1023.
- , 1955, *Bell System Tech. J.*, **34**, 439.
- Gans, R., 1932, *Ann. Physik*, **5**, 15, 28.
- Garwin, R. L., 1957, *I.B.M. J. Res.*, **1**, 304.
- Geusic, Peter, and Schulz-DuBois, 1959, *Bell Sys. Tech. J.*, **38**, 291.
- Gilbert, T. J., and Kelly, J. M., 1955, *Proc. of Conf. on Magnetism and Magnetic Materials*, Pittsburgh (A.I.E.E., N.Y., 1955, p. 253).
- Ginsburg, Richards, and Tinkham, private communication.
- Giordmaine, Alsop, Nash, and Townes, C. H., 1958, *Phys. Rev.*, **109**, 302.
- Goldman, J. E., 1947, *Phys. Rev.*, **72**, 529.
- , 1951, *J. Phys. Rad.*, **12**, 471.

- Goodenough, J. B., 1954, *Phys. Rev.*, **95**, 917.
- , 1955, *Phys. Rev.*, **100**, 564.
- , 1956, *Proc. of Boston Conf. on Magnetism and Magnetic Materials* (A.I.E.E., N.Y., 1957) p. 368.
- , 1958, *J. Phys. Chem. Solids*, **6**, 282.
- Gordon, J. P., and Mowers, K. D., 1958, *Phys. Rev. Lett.*, **1**, 368.
- Graham, C. D., Jr., 1958, *Phys. Rev.*, **112**, 11117.
- , 1959, *Magnetic Properties of Metals and Alloys*, Am. Soc. for Metals, Cleveland.
- Griffiths, Owen, Park, and Partridge, 1957, *Phys. Rev.*, **108**, 1345.
- Griffiths, Owen, Park, and Partridge, 1959, *Proc. Royal Soc.*, **250A**, 84.
- Grimes, D. M., 1957, *J. Phys. Chem. Solids*, **3**, 141.
- Griswold, Kip, and Kittel, 1952, *Phys. Rev.*, **88**, 951.
- Guillaud, C., 1949, *Compt. rend.*, **229**, 818.
- Gyorgy, E. M., 1958, *J. Appl. Phys.*, **29**, 283.
- Hebel and Slichter, 1957, *Phys. Rev.*, **107**, 901.
- Heidenreich, R. D., Nesbitt, E. A., and Burbank, R. D., 1959, *J. Appl. Phys.*, **30**, 995.
- Heine, V., and Pippard, D., 1958, *Phil. Mag.* (8) **3**, 1046.
- Heisenberg, W., 1928, *Z. Physik*, **49**, 619.
- Hofman, Paskin, Tauer, and Weiss, 1956, *J. Phys. Chem. Solids*, **1**, 45.
- Holstein, T., 1951, *Phys. Rev.*, **83**, 1159.
- Holstein, T., and Primakoff, H., 1940, *Phys. Rev.*, **53**, 1098.
- Hu, H., and Weiner, G., 1959, *J. Appl. Phys.*, **30s**, 86s.
- Hughes, Wallace, and Holtzman, 1948, *Phys. Rev.*, **73**, 1277.
- Hutchinson, C. A., and Pastor, R. C., 1951, *Phys. Rev.*, **81**, 282.
- Jacobs, I. S., and Bean, C. P., 1955, *Phys. Rev.*, **100**, 1060.
- , and Luborsky, F. E., 1957, *J. Appl. Phys.*, **28**, 467.
- Jarrett, H. A., Waring, R. K., 1958, *Phys. Rev.*, **111**, 1223.
- Javan, A., 1957, *Phys. Rev.*, **107**, 1579.
- Johnson, C. E., Jr., and Brown, W. F., Jr., 1959, *Phys. Rev.*, **114**, 705.
- Jonker, Wijn, and Braun, 1956, *Philips Tech. Rev.*, **13**, 145.
- Kanamori, J., 1957, *Prog. Theor. Phys., Japan*, **17**, 177, 223.
- Kaplan, H., 1952, *Phys. Rev.*, **85**, 1038.
- , 1959, *Phys. Rev.*, **116**, 888.
- Kastner, T. G., and Kanzig, W., 1957, *J. Phys. Chem. Solids*, **3**, 178.
- Kasuya, T., 1954, *Prog. Theor. Phys.*, **12**, 802.
- Keffer, F., 1955, *Phys. Rev.*, **100**, 1692.
- , and Oguchi, T., 1959, *Phys. Rev.*, **115**, 1428.
- Kelly, J. M., 1956, *Phys. Rev.*, **103**, 499.
- Kersten, M., 1943, *Grundlagen einer neuen Theorie der Ferromagnetischen Hysterese und der Koerzitivkraft* (Hirzel, Leipzig); (Ann Arbor Michigan, Edwards).
- Kip, A. F., Arnold, R. D., 1949, *Phys. Rev.*, **75**, 1556.
- Kip, Kittel, Levy, and Portis, 1953, *Phys. Rev.*, **91**, 1071.
- Kittel, C., 1946, *Phys. Rev.*, **70**, 965.
- , 1947, *Phys. Rev.*, **71**, 270.
- , 1951, *J. de Physique et Rad.*, **12**, 291.
- , 1958A, *Phys. Rev.*, **110**, 836, 1295.
- , 1958B, *Conf. on Low-Temperature Physics*, Leyden.
- , 1959, *Phys. Rev. Letters*, **3**, 169.
- , 1960, *Jour. Appl. Phys.*, **31**, 115.

- , Mitchell, A. H., 1956, *Phys. Rev.*, **101**, 1611.
Kondorsky, E., 1952, *Izvest. Akad. Nauk U.S.S.R., Ser. Fiz.*, **16**, 398.
Koster, G. F., and Slater, H., *Phys. Rev.*, (1959) **113**, 445.
Kouvel, J. S., 1956, *Phys. Rev.*, **102**, 1489.
Kramers, H., 1934, *Physica*, **1**, 182.
Kurti, Robinson, Simon, and Spohr, 1959, *Nature*, **178**, 450.
Landau, L., and Lifshitz, E., 1935, *Physik. Z. Sowjetunton*, **8**, 153.
LeCraw, R. C., and Spencer, E. G., *J. App. Phys.* (to be published).
LeCraw, Spencer, and Porter, 1958, *Phys. Rev.*, **110**, 1311.
LeCraw, Spencer, and Porter, 1958, *J. Appl. Phys.*, **29**, 326.
Lingelbach, R., 1958, *Z. Phys. Chem. (Neue Folge)*, **14**, 1.
Lloyd, J. C., and Smith, R. S., 1959, *J. Appl. Phys.*, **30s**, 274s.
Lomer, W. M., 1960, *Austral. J. Phys.*, **13**, 451.
Luborsky, F. E., Mendelsohn, L. I., and Paine, T. O., 1957, *J. Appl. Phys.*, **28**, 344.
Makhov, Kikuchi, Lambe, and Terhune, 1958, *Phys. Rev.*, **109**, 1399.
———, 1955, *Phys., Rev.*, **97**, 74.
Matthias, B. T., and Corenzwit, 1955, *Phys. Rev.*, **100**, 626.
Matthias, Suhl, and Corenzwit, 1958, *Phys. Rev. Letters*, **1**, 92, 449.
McKeehan, L. W., 1950, *Phys. Rev.*, **79**, 745.
Mitchell, A. H., 1957, *Phys. Rev.*, **105**, 1439.
Morel, P., 1958, *Phys. Rev. Letters*, **1**, 244.
Morin, F. J., 1959, *Bell System Tech. Jour.*, **37**, 1047.
Moriya, Motizuki, Kanamori, and Nagamiya, 1956, *J. Phys. Soc., Japan*, **11**, 211.
Morse, Olsen, and Gavenda, 1959, *Phys. Rev. Letters*, **3**, 15.
Néel, L., 1944, *J. Phys. radium*, **5**, 241, 265.
———, 1946, *Ann. Univ. Grenoble*, **22**, 299.
———, 1947, *Compt. rend.*, **224**, 1488, 1550.
———, 1948, *J. Phys. radium*, **9**, 184, 193.
———, 1949, *Ann. de Geophys.*, **5**, 99.
———, 1950, *J. Phys. radium*, **11**, 49.
———, 1951, *J. Phys. radium*, **12**, 339.
———, 1952, *Ann. Inst. Fourier Grenoble*, **3**, 301.
———, 1953, *Compt. rend.*, **237**, 1468, 1613.
———, 1954, *J. Phys. radium*, **15**, 225.
———, 1956, *J. Phys. radium*, **17**, 250.
———, 1959, *J. Appl. Phys.*, **30s**, 3s.
Nesbet, A., 1958, *Annl. Phys.*, **4**, 87.
Neurath, P. W., 1959, *J. Appl. Phys.* **30s**, 88s.
Newman and Chrenko, 1959, *Bull. Am. Phys. Soc.*, **114**, 53.
Nicholson, R. B., and Nutting, J., 1958, *Phil. Mag.*, **3**, 531.
Overhauser, A. W., 1959, *Phys. Rev. Letters*, **3**, 414.
Overhauser, A. W., 1960, *J. Phys. Chem. Solids*, **13**, 71.
———, and Arrott, A., 1960, *Phys. Rev. Letters*, **4**, 226.
Owen, Browne, Arp, and Kip, 1957, *J. Phys. Chem. Solids*, **2**, 85.
Owen, Browne, Knight, and Kittel, 1956, *Phys. Rev.*, **102**, 1501.
Peter, M., *Phys. Rev.* (to be published).
Peterson, E., and Wrathall, L. R., 1936, *Proc. I.R.E.*, **24**, 275.
Phillips, J. C., 1959, *J. Phys. Chem. Solids*, **11**, 226.
Phillips, J. C., and Kleinman, L., 1959, *Phys. Rev.*, **116**, 287.
Phillips, N. E., 1958, *Phys. Rev. Letters*, **1**, 363.

- Pines, D., 1958, *Phys. Rev.*, **109**, 280.
Pines, D., Schrieffer, J. R., 1958, *Phys. Rev. Letters*, **1**, 407.
Polivanov, K. M., 1952, *Izvest. Akad. Nauk USSR, Ser. Fiz.*, **16**, 449.
Portis, Kip, Kittel, and Brattain, 1953, *Phys. Rev.*, **90**, 988.
Pratt, G., 1955, *Phys. Rev.*, **97**, 926; 1957, **106**, 53.
Pratt and Coelho, 1959, *Bull. Amer. Phys. Soc.*, **114**, 53.
Pry, R. H., and Bean, C. P., 1958, *J. Appl. Phys.*, **29**, 532.
Rado, G. T., and Weertman, J. R., 1954, *Phys. Rev.*, **94**, 1886.
———, Folen, V. J., Emerson, W. H., 1946, *Proc. Inst. Elec. Engrs. (London)*, **104**, Suppl. No. 5, 198.
Rathenau, G. W., 1959, in *Magnetic Properties of Metals and Alloys* (Cleveland, Ohio, American Society for Metals).
Reich, K. H., 1956, *Phys. Rev.*, **101**, 1647.
Reif, F., 1957, *Phys. Rev.*, **106**, 208.
Rickayzen, 1958, *Phys. Rev. Letters*, **2**, 90.
Rodbell, D. S., 1959, *J. Appl. Phys.*, **30**, 187a.
———, and Bean, C. P., 1956, *Phys. Rev.*, **103**, 886.
Sato, H., 1961, *J. Appl. Phys.*, **32**, Suppl. 3, 53.
Sato, Arrott, and Kikuchi, 1959, *J. Phys. Chem. Solids*, **10**, 19.
Schafroth, M. R., 1958, *Phys. Rev.*, **111**, 72.
Schlömann, E., 1958, *J. Phys. Chem. Solids*, **6**, 242; 1958b, **6**, 257.
———, 1959 *Bull. Amer. Phys. Soc., Ser. II*, **4**, No. 1, 53.
Schmitt and Jacobs, 1959, *J. Phys. Chem. Solids*, **3**, 324.
Schrieffer, J. R., (Private Communication).
Schulz-DuBois, E. O., 1959, *Bell Syst. Tech. J.*, **38**, 271.
Seavey, M. H., Tannenwald, P. E., 1956 (May), *Lincoln Lab. Quarterly Prog. Rpt.*
———, ———, 1958, *J. Appl. Phys.*, **29**, 292; 1958b, *Phys. Rev. Letters*, **1**, 168.
Schumacher, R. T., and Slichter, C. P., 1956, *Phys. Rev.*, **101**, 58.
Shirobokov, M., 1939, *Dokl. Acad. Sci. U.R.S.S.*, **24**, 426; 1945, *Zh. eksper. teor. Fiz.*, **15**, 57.
Shull, C. G., and Wilkinson, M. K., 1955, *Phys. Rev.*, **97**, 304.
Shulman, R. G., and Jaccarino, V., 1956, *Phys. Rev.*, **103**, 1126.
Slater, J. C., 1936, *Phys. Rev.*, **49**, 537.
Smart, J. S., 1959, *J. Phys. Chem. Solids*, **11**, 97.
Smith, D. O., 1958, *J. Appl. Phys.*, **29**, 264.
Solt, I. H., White, R. L., Mercereau, J. E., 1958, *J. Appl. Phys.*, **29**, 324.
Spencer, E. G., LeCraw, R. C., Porter, C. S., 1958, *J. Appl. Phys.*, **29**, 429.
Stoner, E. C., and Wohlfarth, E. P., 1948, *Phil. Trans. Roy. Soc. (London)*, **A240**, 599.
Street, R., and Woolley, J. C., 1949, *Proc. Phys. Soc. (London)*, **62A**, 562.
———, and ———, 1950, *Proc. Phys. Soc. (London)*, **62B**, 509.
———, and ———, 1952, *Proc. Phys. Soc. (London)*, **65B**, 679.
Suhl, H., 1956, *Proc. IRE*, **44**, 1270.
———, 1957, *J. Phys. Chem. Solids*, **1**, 209.
Taniguchi, S., and Yamamoto, M., 1954, *Sci. Repts. Res. Inst., Tohoku Univ. Ser. A*, **6**, 330.
Tannenwald, P. E., Seavey, M. H., 1957, *Phys. Rev.*, **105**, 377; *J. Phys. radium* (to be published).
Tiller, C. O., and Clark, G. W., 1958, *Phys. Rev.*, **110**, 583.
Tinkham, M., 1956, *Proc. Roy. Soc. (London)*, **A236**, 535.
———, 1958, *Phys. Rev.*, **110**, 26.

- , paper given at American Physical Society Meeting, N.Y., 1959.
———, and Ferrell, R. A., 1959, *Phys. Rev. Letters*, **3**, 331.
Tinkham, Weinstein, and Kip, 1951, *Phys. Rev.*, **84**, 848.
Tonge, D. G., and Wohlfarth, E. P., 1958, *Phil. Mag.*, **5**, 536.
Tsuya, N., 1958, *J. Appl. Phys.*, **29**, 449.
Van Kranendonk, J., and Van Vleck, J. H., 1958, *Revs. Mod. Phys.*, **30**, 1.
Van Vleck, J. H., *Phys. Rev.*, **52**, 1178.
———, 1950, *Phys. Rev.*, **78**, 266.
———, 1951, *J. Phys. radium*, **12**, 262.
———, 1956, *Phys. Rev.*, **78**, 266.
———, 1959, *J. Phys. radium*, **20**, 124.
Vicena, F., 1955, *Czech. J. Phys.*, **5**, 480.
Wagner, Castle, and Chester, 1959, *Bull. Amer. Phys. Soc.*, **114**, 21.
Walker, L. R., 1957, *Phys. Rev.*, **105**, 390.
Watanabe, H., 1957, *Prog. Theor. Phys.*, **18**, 405.
Weeks, R. A., 1956, *J. Appl. Phys.*, **27**, 1376.
Weertman, J. R., and Rado, G. T., 1958, *J. Appl. Phys.*, **29**, 328.
Weil, L., 1954, *J. Chem. Phys.*, **51**, 715.
Weiss, R. J., and De Marco, J. J., 1958, *Revs. Mod. Phys.*, **30**, 59.
Williams, H. J., and Goertz, M., 1952, *J. Appl. Phys.*, **23**, 316.
———, Shockley, W., and Kittel, C., 1950, *Phys. Rev.*, **80**, 1090.
Wolf, W. P., 1957, *Phys. Rev.*, **108**, 1152.
Wollan, Child, Koehler, and Wilkinson, 1958, *Phys. Rev.*, **112**, 1132.
Wollan, E. O., and Koehler, W. C., 1955, *Phys. Rev.*, **100**, 545.
Wu, Ambler, Hayward, Hoppes, and Hudson, 1957, *Phys. Rev.*, **105**, 1413.
Wwedensky, B., 1921, *Ann. Phys.* (4), **64**, 609.
Yafet, Y., and Kittel, C., 1952, *Phys. Rev.*, **87**, 290.
Yager, W. A., Galt, J. K., Merritt, F. R., 1955, *Phys. Rev.*, **99**, 1203.
———, ———, ———, and Wood, E. A., 1950, *Phys. Rev.*, **80**, 744.
Yosida, K., 1957, *Phys. Rev.*, **106**, 893.
———, and Tachiki, M., 1957, *Prog. Theor. Phys.*, **17**, 331.
Zener, C., 1951, *Phys. Rev.*, **82**, 403.
———, 1951, *Phys. Rev.*, **87**, 440.

Part IV

ELECTRICAL, OPTICAL, AND THERMAL PROPERTIES OF SOLIDS

Chairman:

Prof. ROBERT J. MAUREE
University of Illinois
Urbana, Ill.

Members

Prof. DAVID L. DEXTER
University of Rochester
Rochester, N. Y.

Dr. MELVIN LAX
Bell Telephone Laboratories
Murray Hill, N. J.

Dr. P. G. KLEMENS
Westinghouse Research
Laboratories
Pittsburgh, Pa.

Dr. WALTER J. MERZ
Laboratories RCA Ltd.
Zurich, Switzerland

Dr. J. J. O'DWYER
University of New South
Wales, Australia

Dr. WAYNE W. SCANLON
U.S. Naval Ordnance
Laboratory, White Oak
Silver Spring, Md.

Dr. JAMES H. SCHULMAN
U.S. Naval Research
Laboratory
Washington, D. C.

PREFACE

An adequate survey of the critical problems that limit our understanding of the electrical, optical, and thermal properties of materials is not feasible with the resources available to this panel. A comprehensive review of these areas, which constitute a major fraction of the domain of solid state physics, would fill several volumes. This panel has been forced to adopt a quite different and admittedly less satisfactory procedure.

Seven topics have been selected because of their importance for a fundamental understanding of the interaction of electrons, electro-

magnetic radiation, and lattice vibrations in solids. Each topic has the characteristic that the solution of its problems is of urgent importance for the realization of useful, practical devices. Each topic possesses profound interest for its own sake and is in a lively state of development because its problems, while presenting formidable challenges to the ablest workers, are yielding to systematic inquiry.

It is hoped that the discussion of these seven topics will provide examples of the spirit and techniques with which current materials research in this broad field is being pursued.

DAVID L. DEXTER
P. G. KLEMENS
MELVIN LAX
ROBERT J. MAURER
WALTER J. MERZ
J. J. O'DWYER
WAYNE W. SCANLON
JAMES H. SCHULMAN

CONTENTS

	Page
PREFACE	155
1. ELECTRON AND PHONON PROCESSES IN SOLIDS	158
1.1. Introduction	158
1.2. Electronic Transitions	159
1.2.1. Electromagnetic Interactions with No Magnetic Field ...	159
1.2.2. Electromagnetic Interactions in a Magnetic Field	160
1.3. Phonon Transitions	161
1.4. Electron-Phonon Transitions	163
1.4.1. Photon-Phonon Transitions	163
1.4.2. Transport Phenomena	165
1.5. References	169
2. DIELECTRIC BREAKDOWN IN SOLIDS	171
2.1. Introduction	171
2.2. Dielectric Breakdown of Perfect Crystals	173
2.2.1. General Remarks	173
2.2.2. Low-Temperature Breakdown of Alkali Halides	175
2.2.3. High-Temperature Breakdown of Alkali Halides	177
2.2.4. Non-polar Crystals	178
2.3. Dielectric Breakdown of Glass and Polymers	178
2.4. References	179

3. FERROELECTRICITY	179
3.1. Introduction	179
3.2. State of Knowledge	181
3.2.1. Type of Bonds in Soft Ferroelectrics	182
3.2.2. Type of Bonds in Hard Ferroelectrics	182
3.2.3. Structure and Structure Changes	183
3.2.4. Dipoles and Interaction of Dipoles	183
3.3. Problems Needing Further Research—Some Proposals	185
3.3.1. What Materials are Ferroelectric and Why?	185
3.3.2. Domain Formation and Domain Wall Motion	187
3.4. References	189
4. THERMOELECTRICITY	191
4.1. Introduction	191
4.1.1. The Kelvin Relations	191
4.1.2. Absolute Thermoelectric Power	192
4.2. Thermoelectric Effects in Metals	192
4.3. Thermoelectric Effects in Semiconductors and Insulators	194
4.4. Thermoelectric Conversion	197
4.5. Other Thermoelectric Effects	198
4.6. Conclusions	199
4.7. References	200
5. THERMAL CONDUCTIVITY	200
5.1. Introduction	200
5.2. Electronic Thermal Conductivity of Metals	201
5.3. Lattice Thermal Conductivity of Dielectric Solids	205
5.4. Lattice Thermal Conductivity of Metals	212
5.5. Thermal Conductivity of Superconductors	215
5.6. Thermal Conductivity of Semiconductors	217
5.7. Influence of Other Carriers	219
5.8. References	220
6. ELECTRONIC OPTICAL ABSORPTION	222
6.1. Introduction	222
6.2. Gases	223
6.3. Liquids and Glasses	223
6.4. Solids	224
6.4.1. Centers	224
6.4.2. Pure Solids	229
6.5. References	236
7. LUMINESCENCE	237
7.1. Introduction	237
7.2. Luminescent Centers	238
7.3. Sensitized Luminescence	242
7.4. Luminescence with Charge Transport	244
7.5. Electroluminescence	249
7.6. Organic Luminescent Materials	250
7.7. References	251
7.7.1. Reviews and General Articles	251
7.7.2. Original Articles	252

1. ELECTRON AND PHONON PROCESSES IN SOLIDS

MELVIN LAX

*Bell Telephone Laboratories
Murray Hill, N. J.*

1.1. Introduction

A key factor in the improvement of our understanding of the solid state in the last fifteen years has been the availability of high purity, single crystal material. Thus, well-characterized crystals of high purity and perfection have made possible a study of the *intrinsic* properties of solids. Significant information concerning the *extrinsic* or structure-sensitive properties of these materials has also been obtained by studying the changes in properties which occur when controlled amounts of impurities or imperfections are introduced.

Most of the progress made thus far has been in the study of materials having fairly high mobilities or long mean free paths. For such materials, it is a reasonable idealization to speak of an electronic band system and of a system of lattice oscillators which are only weakly coupled via electron-phonon interactions. The experiments which have been performed and found useful in the study of such materials can therefore be classified in three categories: those involving electronic transitions, vibrational transitions, and combined electronic-vibrational transitions. A comprehensive treatment of phonon-phonon, electron-electron and electron-phonon interactions, as well as the theory of transport phenomena in solids has recently been published by Ziman (1960).

In ionic materials such as ZnO, CdS, and NaCl, where the Coulomb forces produce moderately strong electron-phonon interactions, this classification scheme must be used with some care. For example, an optical transition between the ground state and the excited state of an F-center can be classified as an electronic transition even though many phonons are likely to be emitted simultaneously. This is because such a transition could occur without benefit of phonons. More properly, the energy and oscillator strength corresponding to such a transition are to be regarded as primarily electronic properties, whereas the width of the absorption line is a measure of the electron-phonon interaction. Similar arguments can be applied to cyclotron resonance phenomena. On the other hand, free carrier absorption (a transition between electronic states in the same band) cannot occur without conveying momentum to a phonon, and hence may be

classified as an electron vibrational or "photon-phonon" transition without regard to the strength of the coupling.

There are, of course, materials in which the "conduction band" is so narrow and the effective mass so high, that the estimated mean free path is smaller than the lattice spacing. This point has been emphasized by Joffé (1959) in connection with the behavior of many semiconducting compounds and by Morin (1959) in connection with transition metal oxides. In many such materials, the mobility increases with temperature as $\exp(-\epsilon/kT)$ so that ϵ may be regarded as an activation energy for a jump from one site to another. Under these circumstances it is not proper to speak of bands at all; instead, an electron should be thought of as localized on an atom, interacting with the lattice vibrations to produce a polarization of the lattice in its vicinity. The overlap between electronic wave functions on neighboring atoms is then a weak interaction and conduction occurs only when lattice vibrations increase the overlap.

1.2. Electronic Transitions

1.2.1 Electromagnetic Interactions with No Magnetic Field

One of the more obvious ways of obtaining information about the band structure is to study the onset of direct (i.e., with conservation of quasi-momentum k) absorption between bands. The minimum energy required for strong absorption corresponds to the energy gap between the top of the valence band and the energy of the conduction band at the same k . The rate of increase of absorption with increasing energy in excess of the minimum provides an indication as to whether the transition is allowed or forbidden and hence a clue as to the symmetry of the electronic states involved (Dexter, 1956; Fan, Shephard, and Spitzer, 1956; Bardeen et al., 1956). The magnitude of the absorption cross-section provides a measure of the oscillator strength and through the optical sum rule a rough measure of the effective mass of carriers in the valence band.

In germanium, the valence band at $k = 0$ would be triply degenerate except for spin-orbit splitting. Absorption in the near infrared (Briggs and Fletcher, 1953; Kahn, 1955; Teitler, Burstein and Lax, 1955) displays structure corresponding to transitions between the split off band and the other valence bands, providing direct evidence for a splitting of roughly 0.29 eV.

Additional information concerning the symmetry of the electronic states is sometimes obtained by looking at the optical absorption or edge emission with polarized light. For example,

Hopfield (to be published) has explained the edge emission polarization experiments of Dutton (1958) on CdS, which display an energy shift with direction of polarization, by means of a splitting of the valence band by the crystalline field. Similarly, Callen (1954) has explained the polarization dependent energy gap seen in tellurium under infrared absorption by Loferski (1952) in terms of the valence band structure.

The energy levels associated with shallow impurity states depend on the location of the band edge and the symmetry of the band (Kohn, 1957). For example, donor states in silicon are influenced by the fact that the minimum energy in the conduction band occurs in the $\langle 100 \rangle$ directions about 85% of the distance from $k = 0$ to the zone boundary, leading to a six-ellipsoid picture. The six-ellipsoid picture causes degeneracies among the donor states which are partly lifted by the breakdown of the effective mass approximation in the central cell. Which of the originally degenerate ground state levels are raised, and which are lowered, by the central cell correction can be ascertained by repeating the experiments with a crystal under stress, thus destroying the complete equivalence of the six ellipsoids (Weinreich, Boyle, White and Rodgers, 1959).

1.2.2. Electromagnetic Interactions in a Magnetic Field

Additional information can be obtained by repeating any of the above experiments in a magnetic field (Burstein, et al., 1959; Lax, et al., 1959). The presence of a magnetic field creates "Landau" levels, shifting the energy of free carriers by $(l + \frac{1}{2})\hbar\omega_c$ where $\omega_c = eH/m^*c$ is the cyclotron frequency and m^* is the effective mass. Transitions between various Landau levels of the valence band and various levels of the conduction band permit the determination of $\hbar\omega_c$ and hence the effective mass of carriers in both the valence and conduction bands (Burstein, et al., 1959; Lax, et al., 1959). A plot of the transition energies versus H yields a series of straight lines, one for each choice of l in the two bands. These lines all intersect at $H = 0$ and thus yield a precision determination of the energy gap for the onset of direct absorption (Lax, Roth and Zwerdling, 1959).

Infrared absorption by impurity centers in a magnetic field is a classical Zeeman experiment (Boyle, 1959). The shift, linear or quadratic in H , of the levels serves to insure a proper classification of each level. Since transitions to the various Landau levels can be seen simultaneously in the same experiment, the different energy levels can be measured relative to the continuum, as well

as relative to the ground state. The former (continuum) measurement can be compared directly with effective mass theory, whereas the latter (ground state) involves an unknown central cell correction (Kohn, 1957).

The first experiments to yield direct information about the complicated band structures of silicon and germanium were the cyclotron resonance experiments of Kip and Kittel (1953) and of Lax, et al. (1954). Cyclotron resonance techniques have also been applied successfully to the study of semimetals such as graphite (Galt, et al., 1956) and bismuth (Galt, et al., 1955, 1959) and even of good metals such as zinc (Galt, Merritt, Yager and Dail, 1959). Others have conducted similar experiments on tin, lead, and copper (Fawcett, 1956; Kip, et al., 1957; Bezuglyi and Galkin, 1957).

Cyclotron resonance experiments depend on having relaxation times sufficiently long that $\omega\tau > 1$. Only a few materials are sufficiently pure and have high enough mobility for this condition to be satisfied. The use of infrared magneto-optical absorption methods greatly increases the accessible values of ω , so that a much wider variety of materials can be studied.

Paramagnetic resonance techniques have become powerful tools for the study of electronic structure. We mention in passing that Feher (1959), by using the Electron-Nuclear Double Resonance (ENDOR) technique, was able to observe the hyperfine shift of the resonance line associated with a donor electron. This shift is proportional to $|\psi(r_i)|^2$, the electronic wave function at the nuclear position r_i . In silicon, the interaction is due to the Si^{29} nuclei which are distributed at various positions with respect to the donor. In this way, one obtains a "map" of the donor wave function. Furthermore, $\psi(r)$ contains a sum of contributions from each of the six valleys (Kohn, 1957) with a phase factor $\exp(ik_{0j} \cdot r_i)$ that depends on the location k_{0j} of the minimum in the j -th valley. A detailed analysis has shown that these minima, which were already known by cyclotron resonance to be in the $\langle 100 \rangle$ directions, are actually at 0.85 ± 0.03 of the maximum in these directions.

1.3. Phonon Transitions

Information about the vibration spectra of solids can be inferred from specific heat data. Early attempts to fit the density of states by an expansion into orthogonal polynomials, using the known moments obtained from specific heat data, were not satisfactory because the density of states was shown by van Hove (1953) to have certain inevitable singularities. A moment-singu-

larity analysis suggested by Lax and Lebowitz (1954) was fully developed by Phillips (1959a) and successfully applied to a determination of the density of vibrational states of Ge, Si (Phillips, 1959a), KCl, KBr, NaCl, and NaI (Phillips, 1959b).

For ionic crystals direct information concerning the frequencies of the optical modes of vibration has been obtained from infrared absorption or reflection data. (Mitsuishi, Yoshinaga and Fujita, 1958; Picus, et al., 1959; Collins and Kleinman, to be published).

The most successful method of obtaining direct and fairly complete information about vibration spectra is by inelastic neutron scattering. The change in energy and momentum of the scattered neutron determines the energy and momentum of the emitted phonon. Intensity considerations are used to determine the polarization of the phonon. Extensive studies on silicon and germanium have been completed (Brockhouse and Iyengar, 1958; Brockhouse, 1959; Pelah, Eisenhauer, Hughes and Palevsky, 1957, 1959; Palevsky, et al., 1959). This method can and should be applied to a wide variety of materials. The only limitation on this experiment is the requirement that a large single crystal be available.

It is of interest to note that the suggestion of Lax and Burstein (1955; Lax, 1958) that the infrared lattice absorption spectrum of diamond-type crystals (whose optical modes are inactive because of inversion symmetry) can be explained by means of two-phonon processes has been verified by the neutron determination of the phonon spectrum. Thus, a knowledge of the absorption spectrum yields considerable information about the vibration spectrum, although in a less direct way than neutron scattering. It is likely that, even for the alkali halides whose optical modes permit one-phonon absorption, some of the observed sidebands are associated with two-phonon processes (Lax and Burstein, 1955a; Lax, 1958).

Examination of diffuse X-ray spots (Laval, 1939, 1959; Preston, 1939; Lonsdale, 1943) also yields information about the vibration spectrum (Jacobsen, 1955, 1958). The angle of scattering yields the phonon momentum, although the phonon energies are not measured directly but must be inferred from intensity measurements. This method has been successfully applied to aluminum (Walker, 1956). In general, the intensity in a given direction is made up of contributions from all the vibrational branches, and the original spectrum cannot be deduced directly, but only with the help of a force constant model.

For germanium and silicon, the use of a small number of near-neighbor force constants does not yield an adequate fit with the known experimental spectrum (Lax, 1958). At the very least, long-

range electrostatic forces of dipole-dipole or quadrupole-quadrupole types must be included. The displacement of an atom may not only produce dipole moments in its immediate vicinity, but as a result of dipole-dipole interactions and the polarizability of the lattice, it may induce dipoles much farther away.

1.4. Electron-Phonon Transitions

1.4.1. Photon-Phonon Transitions

When the maximum in the valence band and the minimum in the conduction band do not occur at the same k value, an optical transition with energy equal to the minimum energy cannot occur without the assistance of some mechanism capable of absorbing or emitting the required momentum difference. In a pure crystal, the most likely mechanism involves a phonon, and the process is referred to as an indirect transition (Bardeen, et al., 1956). MacFarlane, McLean, Quarrington and Roberts (1957, 1958, 1959) have examined the same transitions by optical emission techniques. Any phonon having the required k value could contribute, and several such phonons have been identified. In fact, some of the phonons have been seen both in absorption and emission. An average of the photon energy when a phonon is emitted and the energy when a phonon is absorbed yields an accurate value for the indirect gap, whereas the difference between these energies is twice the phonon energy. Agreement is obtained with neutron determinations of the same phonon (Brockhouse and Iyengar, 1958; Brockhouse, 1959) (the latter technique, of course, determines less ambiguously which phonon is being observed). However, with optical techniques, some two-phonon processes appear to have an intensity larger than might be expected, and at least one phonon process is not observed although it does not seem to be excluded on the grounds of symmetry.

The above optical techniques do not require crystals of large size or high purity. For quite impure crystals, however, new "extrinsic" transitions occur. Haynes, et al. (1959) have, for example, observed free holes combining, not with free electrons, but with electrons bound on donor states. Such processes can occur, furthermore, with or without the direct assistance of a phonon. In the latter case the momentum is presumably conveyed to the lattice via the donor nucleus. This technique makes it possible to map out the excited donor states, as well as the ground states, including states that are not accessible by optical absorption from the ground state. Of course, this process is

symmetrical, and the annihilation of free electrons by holes bound on acceptors can be observed.

We have previously remarked that pure optical transitions in which an electron remains in the same band ("free carrier" absorption) are not allowed because energy and momentum conservation cannot be simultaneously obeyed. However, in the radio-frequency or microwave region it is permissible to adopt the old Drude viewpoint (Fan and Becker, 1951) that the "free" carriers move in a "viscous" medium, so that the applied field can do work on the carriers, thus causing absorption of energy. This "viscosity" is of course attributed to collisions with phonons or impurities. A more refined transport theory of free carrier absorption has been given by Herring (1955). The same physical assumptions are involved, but the Herring theory takes into account the energy dependences of the various relaxation processes and the distribution in energy of the free carriers.

In the optical region, and possibly in the infrared as well, it becomes questionable to assume that an electron subject to a rapidly changing electric field engages in dissipative collisions identical to those in the absence of the field. In the high frequency limit, it is indeed possible to demonstrate that the appropriate viewpoint is that originally suggested by Frohlich (Fan and Becker, 1951), and applied in detail to germanium by Meyer (1958) and by Rosenberg and Lax (1958). The absorption can be considered as a two-step process in which a photon is absorbed and a phonon emitted (or absorbed), in either order. Indeed, a quantitative fit at room temperature is obtained for the magnitude and frequency dependence of the absorption. At lower temperatures, however, where impurity scattering begins to replace phonon scattering, the theoretical absorption is not large enough to agree with experiment. There must, furthermore, be a transition from the high frequency, two-step theory, to the low frequency transport theory at frequencies comparable to the lattice vibration frequencies, but this transition has not yet been studied theoretically.

Measurements of the free carrier contribution to the dielectric constant and to the absorption can be used to obtain estimates of an average effective mass and an average relaxation time. Such measurements should be possible in any material with a large enough density of free carriers, even if the relaxation time is too short to permit cyclotron resonance experiments.

It might be interesting to try free carrier absorption experi-

ments at low temperatures in an electric field sufficiently strong to disturb the velocity distribution of the carriers.

It has been observed previously that impurity level transitions can be accompanied by the emission and absorption of phonons. When the coupling is to a flat optical band, the individual lines each involving a particular number of phonons can often be resolved, as in CdS. The relative strength of the individual lines was calculated by Huang and Rhys (1950) for the case of coupling to phonons of a single frequency in terms of modified Bessel functions whose order is the number, p , of the line and whose argument is proportional to the strength of the interaction. For interactions that are moderate, a Poisson distribution prevails, whereas for strong interactions a Gaussian distribution prevails. The moments of the distribution have been given a simple form for an arbitrary strength of interaction (Lax, 1952).

The work of Huang and Rhys has been generalized by Lax (1952) and Lax and Burstein (1955b) to the case of coupling to a set of phonons distributed in frequency. For the case of strong coupling, Lax (1952, 1956) and Lax and Burstein (1955b) have moreover justified the use of the semiclassical Franck-Condon approximation and indicated the cases in which the line shape is expected to be Gaussian.

Radiationless transitions between impurity states or radiationless capture of a "free" electron into an impurity state by simultaneous emission of many phonons is treated by Gummel and Lax (1955, 1957). They show that with the present assumptions concerning the nature of electron-phonon interactions and the usual approximations, such multiphonon processes are highly unlikely in such high mobility materials as silicon and germanium. It appears much more likely that the huge cross sections (10^{-13} cm² to 10^{-12} cm²) observed by Koenig (1958, 1959) for electron capture by donors in germanium are due to a cascade process (Lax, 1959) in which an electron plummets down the excited donor states, emitting a single phonon at each step. The cascade process yields a cross section that varies as $(I/T)/T^4$, roughly in agreement with Koenig's observations both as to magnitude and temperature dependence.

1.4.2. Transport Phenomena

Although transport properties give important information about effective masses and energy band structure (e.g., the multi-ellipsoid conduction band structure of germanium or silicon produces marked anisotropies in the magnetoresistance (Pearson and Suhl,

1950)), they have been also employed to obtain information about electron-phonon interactions. Since there are electron-phonon interactions (Herring, 1955) involving acoustic and optical phonons (in nonpolar as well as polar crystals) of both longitudinal and transverse types, and the interactions may produce intravalley or intervalley collisions, many experiments must usually be performed to obtain definite information about these interactions.

For example, armed with a deformation potential theory and a knowledge of the ellipsoidal band structure in the germanium conduction band, Herring (1955) was able to show that long wavelength acoustical mode scattering is completely characterized by two deformation constants, one appropriate for dilatations and one for shear strains. An analysis of piezoresistance and magnetoresistance measurements at low enough temperatures for other types of scattering to be absent (yet on pure enough materials for impurity scattering not to be important at these temperatures) yielded these two parameters (Herring and Vogt, 1956).

In polar crystals, optical mode scattering is usually dominant (Ehrenreich, 1957, 1959a; Petritz and Scanlon, 1955) in determining the mobility and can be estimated directly from the Coulomb fields produced by the vibrating charged ions. The theory contains only one parameter, the effective charge on the ion, although Ehrenreich (1959b) has shown recently that the presence of an appreciable charge carrier density leads to a modification of the interaction potential as well as the longitudinal optical frequency via screening.

In homopolar crystals, optical mode scattering is less important for mobility and is less well understood. Only recently has it been pointed out (Conwell, 1959; Morgan, 1959) that although optical mode scattering may be relatively unimportant for the mobility (compared to acoustical mode scattering), it still is likely to provide the dominant mechanism for energy loss. The "hot electron" experiments of Morgan (1959) and Gunn (1956, 1959) which examine the reduction in mobility in the presence of a strong electric field are directly sensitive to the energy loss mechanism and yield information concerning optical mode interactions.

Sasaki, Shibuya, Mizuguchi, and Hatoyama (1959) have observed that the mobility in *n*- and *p*-type germanium becomes anisotropic when the field is increased sufficiently. For *n*-type germanium it is clear that electrons in some valleys present a lower effective mass and contribute more current to the total. The effect is apparently enhanced by intervalley scattering from hot to cool valleys. Such anisotropy has been suggested by

Koenig (1958, 1959) as a possible tool in investigating new band structures.

Since the rate at which electrons are captured at impurity sites is influenced by the electron energy distribution, a measure of the decay of excess electrons in the absence or presence of a strong electric field provides a measure of the cross section for capture by thermal or "hot" electrons, respectively. A cascade theory for this capture process is compared by Lax (1959) with experimental results of Koenig (1958, 1959).

The careful work of Geballe, Herring, and Kunzler (1959) on phonon drag effects in thermoelectric power in germanium has given information on phonon-phonon interactions which was completely unavailable before and provided a much better understanding of thermal conduction phenomena.

An excellent example of how the detailed understanding of transport phenomena in germanium and silicon can be applied to other materials is given by Hutson's (1959) study of ZnO. Hutson used Hall effect measurements to obtain a density of states mass of $0.5m$ and a donor ionization energy of 0.05 eV to obtain an effective "polaron" mass of $0.27m$. The temperature dependence of mobility was fitted by a combination of optical and acoustic mode scattering, the former using polaron theory with the polaron mass of $0.27m$. Magnetoresistance measurements provided evidence of at least 12 valleys in the conduction band. Thermoelectric power measurements displayed a phonon drag effect whose saturation effects gave an effective inertial mass of $0.07m$. The small inertial mass is compatible with the density of states mass and with the 12 valleys of small anisotropy, but the donor binding energy and optical mode scattering seem to require an effective mass four times larger.

In view of the excellent summary of recent developments in transport phenomena in semiconductors available from the proceedings of the 1958 Rochester Conference, this section will be concluded with remarks on experiments that have been found useful in studying metals.

The "classical" methods of magnetoresistance (Chambers, 1960) and de Haas-van Alphen oscillations (Shoenberg, 1960) have been most useful. Shoenberg (1939, 1952) has for example been able to arrive at a qualitatively correct picture of the enormously complicated bismuth band structure on the basis of such methods.

Kunzler and co-workers (Kunzler and Hsu, 1960; Boyle, Hsu and Kunzler, 1960) have recently discovered an interesting effect in bismuth which is closely related to the de Haas-van Alphen

oscillations. They found an orientation-dependent oscillatory variation of temperature with reciprocal magnetic field, from which effective masses may be deduced. They suggested that this new technique may provide a simpler method of observation than de Haas-van Alphen experiments.

Pippard (1959) and Chambers (1956) have demonstrated that by studying the anomalous skin effect (Smith, 1960) particularly in a magnetic field, information about the location of the Fermi surface in a metal can be obtained. Pippard (1959) has in this way determined in detail the shape of the Fermi surface in copper. Pippard's computational procedure involves perturbation from a spherical shape, with a certain amount of trial and error. It is not clear how successful such a procedure would be in studying really complicated Fermi surfaces like those in bismuth or graphite. For the latter, the cyclotron resonance procedure suggested by Azbel' and Kaner (1956) and exploited by Galt, et al., (1955, 1956, 1959) and Lax (1958) seems to be preferable.

Of course the cyclotron resonance procedure depends on having a sufficiently long mean free path, i.e., a crystal with sufficiently low residual resistance (Kip, 1960). This restriction can be relaxed by working with large magnetic fields and by using infrared or optical rather than microwave techniques.

Morse (1960) and co-workers at Brown University have also developed what is likely to be an unusually useful technique: magnetic oscillations of ultrasonic attenuation in metals. This phenomenon involves several characteristic lengths, among which are the wavelength of sound, the electron mean free path and the cyclotron orbit radius, and is likely to yield additional information if the experiments cover several ranges of frequency, magnetic field, and temperature. The disentanglement of the experimental information will depend heavily on a detailed theory. Fortunately contributions to the theory have been made by Pippard (1951, 1960) and Rodriguez (1958), and other recent work has been done by Kjeldaas and Holstein (1959) and by Cohen and Heine (1958).

For recent attempts to compare experimental results using various techniques with energy band theory of the alkali and noble metals see the review by Cohen and Heine (1958) and the report of the conference on the Fermi surface edited by Harrison and Webb (1960).

1.5. References

- Azbel', M. Ya., and Kaner, E. A., 1956, *Zhur Eksptl. i. Teoret. Fiz.*, **30**, 811 (translation: *Soviet Phys. JETP*, 1956, **3**, 772).
- , 1958, *J. Phys. Chem. Solids*, **8**, 318.
- Bardeen, J., Blatt, F. J., and Hall, L. H., 1956, in *Photoconductivity Conference* (John Wiley and Sons, New York).
- Bezuglyi, P. A., and Galkin, A. A., 1957, *Zhur Eksptl. i. Teoret. Fiz.*, **33**, 1076 (translation: *Soviet Phys. JETP*, 1958, **6**, 831).
- Boyle, W. S., 1952, *J. Phys. Chem. Solids*, **3**, 318.
- , Hsu, F. S. L., and Kunzler, J. E., 1960, *Phys. Rev. Letters*, **4**, 278.
- Breckenridge, R. G., *et al.*, editors, 1956, *Photoconductivity Conference* (John Wiley and Sons, New York).
- Briggs, H. B., and Fletcher, R. C., 1953, *Phys. Rev.*, **91**, 1342.
- Brockhouse, B. N., 1959, *Phys. Rev. Letters* **2**, 256.
- , and Iyengar, P. K., 1958, *Phys. Rev.*, **111**, 747.
- Burstein, E., Picus, G. S., Wallis, R. F., and Blatt, F. J., 1959a, *J. Phys. Chem. Solids*, **8**, 305.
- , 1959b, *Phys. Rev.*, **113**, 15.
- Callen, H. B., 1954, *J. Chem. Phys.*, **22**, 518.
- Chambers, R. G., 1956, *Proc. Roy. Soc.*, **A238**, 344.
- , 1960, in *The Fermi Surface*, Harrison, W. A., and Webb, M. B., eds. (John Wiley and Sons, New York), pp. 100–124.
- Cohen, M. H., and Heine, V., 1958, *Phil. Mag. Suppl.* **2**, 395.
- Collins, R. J., and Kleinman, J. J., *J. Phys. Chem. Solids* (to be published).
- Conwell, E., 1959, *J. Phys. Chem. Solids*, **8**, 234.
- Dexter, D. L., 1956, in *Photoconductivity Conference* (John Wiley and Sons, New York).
- Dexter, R. N., Lax, B., and Zeiger, H. J., 1954, *Phys. Rev.*, **95**, 557.
- Dresselhaus, G., Kip, A. F., and Kittel, C., 1953, *Phys. Rev.*, **22**, 827.
- Dutton, D., 1958, *J. Phys. Chem. Solids*, **6**, 101.
- Ehrenreich, H., 1957, *J. Phys. Chem. Solids*, **2**, 131; *ibid.*, 1959a, **8**, 130; *ibid.*, 1959b, **9**, 129.
- Fan, H. Y., and Becker, M., 1951, in *Proceedings of the Reading Conference* (Butterworths Scientific Publications, Ltd., London), pp. 132–147.
- Fan, H. Y., Shephard, M. L., and Spitzer, W., 1956, in *Photoconductivity Conference* (John Wiley and Sons, New York).
- Fawcett, E., 1956, *Phys. Rev.*, **103**, 1582.
- Feher, G., 1959, *J. Phys. Chem. Solids*, **8**, 486.
- Galt, J. K., Merritt, F. R., Yager, W. A., and Dail, H. W., 1959, *Phys. Rev. Letters*, **2**, 292.
- Galt, J. K., Yager, W. A., and Dail, H. W., 1956, *Phys. Rev.*, **103**, 1586.
- Galt, J. K., Yager, W. A., Merritt, F. R., Cetlin, B. B., and Dail, H. W., 1955, *Phys. Rev.*, **100**, 748.
- Galt, J. K., Yager, W. A., Merritt, F. R., Cetlin, B. B., and Brailsford, A. P., 1958, *Phys. Rev.*, **114**, 1396.
- Geballe, T. H., Herring, C., and Kunzler, J. E., 1959, *J. Phys. Chem. Solids*, **8**, 347.
- Gummel, H., and Lax, M., 1955, *Phys. Rev.*, **37**, 1469.
- , and ———, 1957, *Ann. Phys.*, **2**, 28.
- Gunn, J. B., 1956, *J. Electronics*, **2**, 87.
- , 1959, *J. Phys. Chem. Solids*, **8**, 239.

- Harrison, W. A., and Webb, M. B., 1960, eds., *The Fermi Surface* (John Wiley and Sons, New York).
- Haynes, J. R., Lax, M., and Flood, W. F., 1959, *J. Phys. Chem. Solids*, **8**, 392.
- Herring, C., 1955, *Bell Systems Tech. J.*, **34**, 237.
- Herring, C., and Vogt, E., 1956, *Phys. Rev.*, **101**, 944.
- Hopfield, J., *J. Phys. Chem. Solids* (to be published).
- Hove, van L., 1953, *Phys. Rev.*, **89**, 1189.
- Huang, K., and Rhys, A., 1950, *Proc. Roy. Soc. (London)*, **A204**, 406.
- Hutson, A. R., 1959, *J. Phys. Chem. Solids*, **8**, 467.
- Jacobsen, E. H., 1955, *Phys. Rev.*, **97**, 654.
- , 1958, *Revs. Mod. Phys.*, **30**, 234.
- Joffé, A., 1959, *J. Phys. Chem. Solids*, **8**, 2.
- Kahn, A. H., 1955, *Phys. Rev.*, **97**, 1647.
- Kip, A. F., 1960, in *The Fermi Surface* (John Wiley and Sons, New York), pp. 146–153.
- Kip, A. F., Langenberg, D. N., Rosenblum, B., and Wagoner, G., 1957, *Phys. Rev.*, **108**, 494.
- Kjeldaas, T., and Holstein, T., 1959, *Phys. Rev. Letters*, **2**, 340.
- Koenig, S., 1958, *Phys. Rev.*, **110**, 986.
- , 1959, *J. Phys. Chem. Solids*, **8**, 227.
- Kohn, W., 1957, in *Solid State Physics* (Academic Press, New York), Vol. 5.
- Kunzler, J. E., and Hsu, F. S. L., 1960, in *The Fermi Surface* (John Wiley and Sons, New York), p. 38.
- Laval, J., 1939, *Compt. rend.*, **308**, 1512.
- , 1959, *Revs. Mod. Phys.*, **30**, 229.
- Lax, B., Roth, L. M., and Zwerdling, S., 1959, *J. Phys. Chem. Solids*, **8**, 311.
- Lax, B., Zeiger, H. J., Dexter, R. N., and Rosenblum, E. S., 1954, *Phys. Rev.*, **93**, 1418.
- Lax, M., 1952, *J. Chem. Phys.*, **20**, 1752.
- , 1956, "The Influence of Lattice Vibrations on Electronic Transitions in Solids," in *Photoconductivity Conference* (John Wiley and Sons, New York).
- , 1958, *Phys. Rev. Letters*, **1**, 131; *ibid*, **1**, 133.
- , 1959, *J. Phys. Chem. Solids*, **8**, 66.
- , and Burstein, E., 1955a, *Phys. Rev.*, **97**, 69; *ibid*, 1955b, **100**, 592.
- , and Lebowitz, J. L., 1954, *Phys. Rev.*, **96**, 594.
- Loferski, J., 1952, *Phys. Rev.*, **87**, 905.
- , 1953, *Thesis*, University of Pennsylvania.
- Mac Farlane, G. G., McLean, T. P., Quarrington, J. E., and Roberts, V., 1957, *Phys. Rev.*, **108**, 1377; 1958, *ibid*, **111**, 1245.
- , 1959, *J. Phys. Chem. Solids*, **8**, 388.
- Meyer, H. J. G., 1958, *Phys. Rev.*, **112**, 298.
- Mitsubishi, A., Yoshinaga, H., and Fujita, S., 1958, *J. Phys. Soc. (Japan)*, **13**, 1235.
- Morgan, T. N., 1959, *J. Phys. Chem. Solids*, **8**, 245.
- Morin, F. J., 1959, *Bull. Amer. Phys. Soc. II*, **4**, 186.
- Morse, R. W., 1960, in *The Fermi Surface* (John Wiley and Sons, New York), pp. 214–223.
- Palevsky, H., Hughes, D. J., Kley, W., and Tunkelo, E., 1959, *Phys. Rev. Letters*, **2**, 258.
- Pearson, G. L., and Suhl, H., 1950, *Phys. Rev.*, **83**, 768.

- Pelah, I., Eisenhauer, C. M., Hughes, D. J., and Palevsky, H., 1957, *Phys. Rev.*, **108**, 1081; 1959, *ibid*, **118**, 49.
- Petriz, R. L., and Scanlon, W. W., 1955, *Phys. Rev.*, **97**, 1620.
- Phillips, J. C., 1959a, *Phys. Rev.*, **113**, 147; 1959b, *Bull. Amer. Phys. Soc. II*, **4**, 142.
- Picus, G. S., Burstein, E., Hennis, B. W., and Hass, M., 1959, *J. Phys. Chem. Solids*, **8**, 282.
- Pippard, A. B., 1957, *Phil. Trans. Roy. Soc. (London)*, **A250**, 325.
- , 1957, *Phil. Mag.*, **2**, 1147.
- , 1960, in *The Fermi Surface* (John Wiley and Sons, New York), pp. 224–232.
- Preston, G. D., 1939, *Proc. Roy. Soc. (London)*, **A172**, 116.
- Rodriguez, S., 1958, *Phys. Rev.*, **112**, 80.
- Rosenberg, R., and Lax, M., 1958, *Phys. Rev.*, **112**, 843.
- Sasaki, W., Shibuya, M., Mizuguchi, K., and Hatoyama, G. M., 1959, *J. Phys. Chem. Solids*, **8**, 250.
- Shoenberg, D., 1939, *Proc. Roy. Soc. (London)*, **A170**, 341.
- , 1952, *Phil. Trans. Roy. Soc.*, **A245**, 1.
- , 1960, in *The Fermi Surface*, Harrison, W. A., and Webb, M. B., eds. (John Wiley and Sons, New York), pp. 74–84.
- Smith, G. E., 1960, in *The Fermi Surface* (John Wiley and Sons, New York), pp. 182–196.
- Teitler, S., Burstein, E., and Lax, M., 1955, *Phys. Rev.*, **98**, 1561.
- Walker, C. B., 1956, *Phys. Rev.*, **103**, 547, 558.
- Weinreich, G., Boyle, W. S., White, H. G., and Rodgers, 1959, *Phys. Rev. Letters*, **2**, 96.
- Ziman, J. M., 1960, *Electrons and Phonons* (Clarendon Press, Oxford).

2. DIELECTRIC BREAKDOWN IN SOLIDS

J. J. O'DWYER

*University of New South Wales
Australia*

2.1. Introduction

Dielectric breakdown under engineering conditions differs from that obtained under controlled laboratory conditions. In the former a dielectric may be destroyed by the violence of the discharge in the ambient medium, by discharges resulting from chemical contamination of the surface, by macroscopic physical defects in the dielectric itself, or by slow chemical changes taking place in the dielectric which alter the breakdown strength. The former type of breakdown can be eliminated in laboratory work by suitable selection of the ambient medium, the electrode configuration, and the dielectric material itself, and it is with the latter type of breakdown that this article is exclusively concerned.

The aim of work on the dielectric breakdown of solids may either be a detailed physical understanding of the process or de-

developmental research requiring mainly a phenomenological theory. The principal substances which have been studied for their academic interest are the alkali halides, whereas from a technical point of view glass and the various polymers are important. The main subdivisions of this article will therefore be perfect crystals (mainly the alkali halides) on the one hand and glass and polymers on the other. Before proceeding to the discussion of special substances, it is important to distinguish among the various breakdown criteria which have been proposed for different substances in different temperature ranges. These are the intrinsic, thermal, and avalanche criteria of breakdown.

The notion of intrinsic breakdown was developed by von Hippel (1935) and Frohlich (1937, 1947, 1956), and it refers to a class of breakdown for which no equilibrium is possible between the rate at which conduction electrons receive energy from the electric field and the rate at which they transfer energy to the lattice. Conduction electrons receive energy from an electric field F at a rate

$$A = jF \quad (2.1)$$

where j is the electric current density. If the rate at which some particular process transfers energy from the conduction electrons to the lattice be denoted by B , then the condition for a steady state is

$$A = B \quad (2.2)$$

In general, condition (2.2) cannot be satisfied for values of F above some critical value F^* . This is the critical field strength for the particular transfer process which is being considered. The physical problem of intrinsic breakdown is thus to determine details of that energy transfer process from conduction electrons to lattice which is the ultimate in this sense: When it cannot supply a steady state, no other process remains which can do so before the dielectric fails. In cases of intrinsic breakdown the destruction of the lattice is of course thermal, but the condition for the initial irreversible instability is determined from steady state electronic processes.

The fundamental equation governing thermal breakdown is

$$C_v \frac{dT}{dt} - \text{div} (K \text{ grad } T) = \sigma F^2, \quad (2.3)$$

where C_v is the specific heat per unit volume, K and σ are the thermal and electrical conductivities respectively, T is the temperature, and t is the time. This relation simply states that heat sup-

plied either raises the temperature or is conducted away. If satisfaction of this equation involves a higher temperature than the dielectric can bear, thermal breakdown results. This type of breakdown is of significance in practical applications and in such a context it is treated extensively by Whitehead (1950). Two special cases arise, the first of which is for extremely slow time variation of all physical quantities and leads to

$$\text{div} (K \text{ grad } T) + \sigma F^2 = 0, \quad (2.4)$$

which may be said to determine the steady state thermal breakdown strength. The second is for the case in which the applied voltage rises so quickly that thermal conduction processes are unimportant and is described by writing

$$C_v \frac{dT}{dt} = \sigma F^2, \quad (2.5)$$

which may be said to determine the "impulse" thermal breakdown strength. In this latter case, the breakdown is always a function of the rate of application of the voltage (Franz, 1956) (Vermeer, 1956). It is also the case which is normally of most interest under laboratory conditions.

A criterion for avalanche breakdown was given by Seitz (1949) which consists essentially in estimating the critical size of an avalanche produced by successive internal ionizations, and then calculating the field strength required to produce this avalanche. This approach is related to intrinsic breakdown but introduces a distinctly statistical element into the description of the breakdown process, since it considers the process not so much as the failure of some steady state mechanism but as one "successful" event out of many "unsuccessful" ones. This criterion would seem to be most suitable for application to a dielectric in which the conduction electron density is very low for all conditions of temperature and field strength.

All of these concepts of the mechanism of breakdown have been used with more or less success for different materials under different physical conditions. The purpose of this article is to examine the application of these criteria with particular emphasis on situations in which the evidence for such application is doubtful or incomplete.

2.2. Dielectric Breakdown of Perfect Crystals

2.2.1. General Remarks

Among perfect crystals, the suitability of the alkali halides for what may be regarded as the academic study of dielectric break-

down has its roots ultimately in their simple ionic crystal structure. This has resulted in a theoretical understanding of the physical properties of alkali halides which is in a state at least as advanced as that pertaining to any other group of solids, and certainly more advanced than that pertaining to any other group of dielectric solids.

The experimental work on alkali halides has fallen into two classes. On the one hand, the breakdown strength of a particular substance has been measured as a function of temperature, and, on the other, breakdown strengths at a particular temperature have been measured for various substances. The latter work has proved strangely unfruitful since almost any theory of breakdown will give the correct manner of variation from substance to substance within the alkali halides. It ought also be mentioned in this connection that work of Frohlich and Paranjape (1956) gives theoretical reasons for supposing that measurements at fixed fraction of the Debye temperature for each substance are more relevant than comparison of measurements at fixed temperature. For these reasons the breakdown strength as a function of temperature is regarded as more significant.

A study of experimental papers on the breakdown of alkali halides shows great diversity of results, but two conclusions emerge fairly strongly (O'Dwyer, 1958). They are:

1. The breakdown strength is of the order of 10^6 volts per cm and increases very slowly with temperature from very low temperatures up to about room temperature. Within this temperature range breakdown is believed to be intrinsic.
2. In the high temperature region of breakdown, which extends roughly from room temperature upwards, there is great conflict of experimental opinion, much of which would be resolved if breakdown is found to be of thermal nature in this region.

Therefore the high- and low-temperature regions of breakdown in alkali halides will be discussed separately in the following section.

Besides the alkali halides, other crystalline dielectrics which might be mentioned are germanium, silicon, diamond, crystalline quartz, and mica. The first two of these are normally regarded as semiconductors and very little work has been done on diamond and crystalline quartz, but mica is of interest especially since there is some evidence that it may follow an avalanche breakdown mechanism.

2.2.2. Low-Temperature Breakdown of Alkali Halides

Since thermal conductivity decreases and electrical conductivity increases with increasing temperature, thermal breakdown at low temperatures is unlikely and intrinsic breakdown is probable, and experimental evidence confirms this. Three distinct theories of intrinsic breakdown of alkali halides in the low-temperature region have been proposed. The two early theories are both single electron approximations in that they base their criteria for breakdown on the analysis of the average behavior of a single conduction electron. The first of these, due to von Hippel (1935) and Callen (1949), conceives the critical field strength as that which is capable of producing a net acceleration of every conduction electron against the retarding influence of inelastic collisions with the lattice vibrations. This has been called the "low-energy criterion" since it is the low-energy electrons which encounter most friction with the lattice. Generally speaking it gives results for the critical field strength which are above the observed breakdown strengths. The other single-electron theory is that of Frohlich (1937, 1939) in which the critical field strength is that which produces a net acceleration of the average conduction electron whose energy is at least equal to the thermal ionization energy. It is easily shown that, at this critical field strength, internal ionization processes cannot be balanced by the inverse process of recombination. Since the internal ionization energy is high compared with thermal energies this theory is referred to as the "high energy criterion"; it gives results for the critical field strength somewhat lower than the observed breakdown strengths.

A more sophisticated theory which includes the effect of interactions among conduction electrons has been given by Frohlich and Paranjape (1956). This theory is applicable when the density of conduction electrons at the high field strengths near breakdown exceeds a certain critical density, above which they exchange energy among themselves more rapidly than with the lattice. The energy distribution of the conduction electrons is then Maxwellian with some temperature T larger than the lattice temperature T_0 , but not orders of magnitude greater. It is in this sense that electron-electron interaction has been included.

This Maxwellian distribution of conduction electrons is essential to the Frohlich-Paranjape theory, and estimates have been made of the density required to secure it. For temperatures so low that most of the thermal electrons cannot, from energy considerations, emit an optical quantum (i.e., for lattice temperatures

very much lower than the Debye temperature), the energy transfer will arise from the very few high energy electrons or from scattering by non-polar acoustical modes. In these circumstances very low electron densities should make the energy distribution Maxwellian, so that from this point of view the low temperature region seems to be that in which to seek verification of the Frohlich-Paranjape theory. It is interesting to note that Stratton (1958) has shown that, for polar crystals at low temperatures and low field strengths, the conduction electron mobility should be strongly dependent on whether electron-electron or electron-lattice collisions predominate. However, for the alkali halides, very low temperatures (liquid helium) or very high field strengths (near breakdown) may be required to throw some light on the electron distribution function. A further point worth noting is that the failure of the steady state corresponding to the "high energy criterion" critical field strength could be responsible for producing the physical situation in which a collective theory would hold; this could only be so if the single-electron critical field strength were less than the collective critical field strength.

However, apart from doubt regarding the validity of the model, a further difficulty arises in the Frohlich-Paranjape theory of breakdown of polar crystals. This difficulty is a theoretical one, namely, that the use of perturbation theory to describe the interaction between a conduction electron and lattice vibrations is satisfactory only if that interaction is small. A conduction electron with energy of the order of several optical quanta interacts very strongly with a polar lattice, and it can, in fact, be shown that perturbation theory does not validly treat this problem (Frohlich, 1954). Since the electron temperature at the critical field strength is given by

$$T^* \approx \hbar\omega/k, \quad (2.6)$$

where $\hbar\omega$ is the optical lattice quantum energy and k is the Boltzmann constant, most of the conduction electrons will have energy in the range for which perturbation theory is invalid, and the final result is therefore suspect on theoretical grounds. Frohlich and Paranjape endeavor to overcome this difficulty to some extent by expressing the critical field strength in terms of another measured constant of the material which is also itself calculated by perturbation theory. In this way it is hoped that a repeated unjustified use of perturbation theory (but worked backwards so to speak) may give a more nearly correct result than one such use. Thus, the result for the critical field strength is expressed

in terms of the low field strength electronic mobility, giving (Frohlich and Paranjape, 1956)

$$F^* \approx 0.5 \frac{(\hbar\omega)^{1/2}}{2m} \frac{1}{\mu(T_0) n(T_0)}, \quad (2.7)$$

where $n(T_0)$ is the number of quanta at temperature T_0 , and $\mu(T_0)$ is the electronic mobility at that temperature. Comparison of the somewhat dubious experimental figures available for F^* and μ shows that Eq. (2.7) relates them correctly at least to an order of magnitude.

A summary of the outstanding difficulties in the low temperature region of breakdown of the alkali halides would thus appear to be:

1. Granted that experimental evidence points to the intrinsic nature of breakdown in this region, the experimental concept of the breakdown strength of a pure perfect crystal could be made more precise by sorting out the effects due to crystalline defects. This has been done to some extent by Cooper and Wallace (1954), who have demonstrated that annealing can greatly influence the breakdown strength.

2. On the basis of the present theory, the main theoretical difficulty is an adequate treatment of the motion of an electron in a polar lattice in the case of very strong interaction.

2.2.3. High-Temperature Breakdown of Alkali Halides

In the high-temperature region experimental evidence as to whether breakdown is intrinsic or thermal is conflicting. However, it seems possible that most of the experimental work (if correctly interpreted) tends rather to favor an impulse thermal theory of breakdown. Thus, the successive application of pulses of gradually increasing maximum voltage, as done by Calderwood and Cooper (1953), may not differ greatly from a "d.c." test as far as thermal breakdown is concerned, especially if the heat capacity of the electrode system is small. For true comparison, both the electrical and thermal conditions should be the same if thermal breakdown is suspected.

Frohlich (1947) and O'Dwyer (1957) have proposed theories of intrinsic breakdown in which the principal mechanism by which electrons transfer energy to the lattice is by transitions from conduction levels to isolated levels involving the emission of lattice quanta. A theory of impulse thermal breakdown has been developed by Franz (1952) and Veelken (1954) and applied to the alkali halides. Their field strength has a negligible temperature variation which is not in accordance with the experimental facts.

However, they consider electronic conductivity only, which seems quite unjustified in substances such as the alkali halides; there appears more justification for considering only ionic conductivity.

It appears that the true nature of breakdown in the high-temperature region and a reasonable theory remain to be determined.

2.2.4. Non-Polar Crystals

The situation for nonpolar insulating crystals is different from that for the alkali halides. In the first place, there is almost no experimental work available, perhaps for the reason that it is hard to think of any simple nonpolar crystal (other than diamond) which is insulating at room temperature. In this connection measurements on germanium and silicon at liquid helium temperatures may be useful.

On the theoretical side, however, this case is most interesting since here the Frohlich-Paranjape theory is not using perturbation theory outside its range of validity. The remarkably simple result for the breakdown field strength for temperatures well below the Debye temperature is

$$F^* \approx 0.17 \frac{\hbar N_a^{1/3}}{m \mu(\theta)} \quad (2.8)$$

where N_a is the number of atoms per unit volume, and $\mu(\theta)$ the electronic mobility at the Debye temperature in a low field strength. A suitably conceived experimental program for non-polar insulating crystals (perhaps at very low temperatures as indicated above) would thus be very interesting in order to check the Frohlich-Paranjape theory for these substances for which it should be most reliable.

2.3. Dielectric Breakdown of Glass and Polymers

Generally speaking, glass and polymers show similar breakdown characteristics. For sufficiently low temperatures, the breakdown strength is of the order of 10 megavolts per cm, varies slowly with temperature, and is probably intrinsic. For higher temperatures or long durations of the applied voltage the breakdown strength decreases with increasing temperature.

The dielectric breakdown of certain types of glass has been thoroughly investigated by Keller (1948, 1952) and Vermeer (1954, 1956, 1959). They have proved that the breakdown is of the "impulse thermal" type for glass in the high-temperature region under their experimental conditions. Vermeer calculated the impulse thermal breakdown strength from Eq. (2.5) using

experimentally measured values of the electrical conductivity and obtained good agreement with measured breakdown strengths. The only outstanding problem in this region would be a theory of the field- and temperature-dependence of the electrical conductivity. In the low temperature region the mechanism of the breakdown of glass is not known, and there is consequently no theory.

There has not been a great deal of work on the breakdown of polymers apart from "d.c" measurements of the temperature variation of the breakdown strength (Oakes, 1948, 1949). However, since the ratio of the thermal to the electrical conductivity is greater than that for other insulators by many powers of ten, it would seem to be reasonable to seek some form of intrinsic or avalanche breakdown in these substances.

2.4. References

- Callen, H. B., 1949, *Phys. Rev.*, **76**, 1394.
 Cooper, R., and Wallace, A. A., 1953, *Proc. Phys. Soc. (London)*, **B66**, 1113.
 Franz, W., 1952, *Z. Physik*, **132**, 285.
 ———, 1956, *Encyclopedia of Physics* (Springer-Verlag, Berlin), Vol. 17, p. 155.
 Frohlich, H., 1937, *Proc. Roy. Soc.*, **A160**, 230; 1939, *ibid*, **172**, 94; 1947, *ibid*, **188**, 521.
 ———, 1954, *Phil. Mag. Suppl.*, **3**, 325.
 ———, and Paranjape, B. V., 1956, *Proc. Phys. Soc. (London)*, **B69**, 21.
 Hippel, von A. R., 1935, *Ergebn. exakt. Naturw.*, **14**, 79.
 Keller, K. J., 1948, *Physica*, **14**, 475; 1951, *ibid*, **17**, 511.
 Oakes, W. G., 1948, *J. Brit. Inst. Radio Engrs.*, **95**, I, 36; 1949, *ibid*, **96**, I, 37.
 O'Dwyer, J. J., 1957, *Proc. Phys. Soc. (London)*, **B70**, 761.
 ———, 1958, *Phil. Mag. Suppl.*, **7**, 340.
 Seitz, F., 1949, *Phys. Rev.*, **76**, 1376.
 Stratton, R., 1958, *Proc. Roy. Soc.*, **A246**, 406.
 Veelken, R., 1955, *Z. Physik*, **142**, 476, 544.
 Vermeer, J., 1954, *Physica*, **20**, 313; 1956, *ibid*, **22**, 1247, 1257, 1269.
 Whitehead, S., 1950, *Dielectric Breakdown of Solids* (Oxford University Press, Oxford).

3. FERROELECTRICITY

WALTER J. MERZ
Laboratories RCA Ltd.
Zurich, Switzerland

3.1. Introduction

Valasek (1921) first pointed out the analogy between the dielectric properties of Rochelle salt (RS) and the magnetic properties of a ferromagnetic material. Thus Rochelle salt, $\text{KNaC}_4\text{H}_4\text{O}_6 \cdot 4\text{H}_2\text{O}$

was the first material discovered to be ferroelectric. Although a number of careful investigations were performed during the following years, not much fundamental information about ferroelectricity could be obtained; RS was a much too complicated material. A significant discovery was that the deuterated Rochelle salt is ferroelectric, too; Hablützel (1939) found for the deuterated salt a shift of the upper Curie point from 23°C to 35°C and of the lower Curie point from -18°C to -23°C.

Busch and Scherrer (1935) discovered a second class of ferroelectrics, the class typified by KH_2PO_4 . Since this material has a much simpler chemical formula and crystallographic structure, there was great hope of understanding the fundamentals of ferroelectricity. In the course of about 10 years, a number of crystals of the composition XH_2YO_4 with X either K, Rb, or Cs, and Y either P or As were found to be ferroelectric as long as they crystallize with the symmetry V_d (Matthias, Merz, and Scherrer, 1947; Matthias, 1952; Frazer, 1953). The Curie points of all of these materials occur at very low temperatures. The corresponding deuterated salts are also ferroelectric and, in contrast to RS show a very pronounced isotope effect. The upwards shift of the Curie point is of the order of 60–90°C for the different materials. This first-order isotope effect and the fact that up to this point all ferroelectrics contained hydrogen, led at that time to the conclusion that ferroelectricity must be caused by O-H-O hydrogen bonds.

This assumption was soon shown to be incorrect. During World War II, von Hippel (1950) in the USA and Wul (1946) in the USSR discovered the anomalous dielectric properties of BaTiO_3 , a material containing no hydrogen bonds. The discovery of ferroelectricity in BaTiO_3 was a great stimulus to the development of this field of solid-state physics, because the perovskite structure in which BaTiO_3 crystallizes is a simple one. After World War II a number of other materials with perovskite structure were found to be ferroelectric. For example, Matthias and Remeika (1951) discovered the new ferroelectrics KNbO_3 , KTaO_3 , and others and Shirane et al. (1951) discovered PbTiO_3 . The main feature of the perovskite structure is the presence of oxygen octahedra each containing a positive ion in its center (Ti, Nb, Ta). It was therefore a natural step to investigate other crystal classes containing oxygen octahedra. Cook and Jaffe (1952) discovered the ferroelectric properties of $\text{Cd}_2\text{Nb}_2\text{O}_7$, a material which crystallizes in the pyrochlore structure and contains oxygen octahedra.

The discovery of guanidine aluminum sulfate hexahydrate (GASH), $(\text{CN}_3\text{H}_3)\text{Al}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$, in 1955 as a new ferroelectric

by Holden, Matthias, Merz, and Remeika (1953, 1956), however, turned out to be the major step in the discovery of new ferroelectrics. It became clear that many compounds with hydrogen bonds O-H-O and O-H-N are ferroelectric, especially when these bonds are linking groups like $(\text{SO}_4)^{--}$, $(\text{SeO}_4)^{--}$ etc. Matthias and Pepinsky and others have discovered many new ferroelectrics during the last few years; some of them are: alums (Pepinsky, et al., 1956); ammonium sulfate, $(\text{NH}_4)_2\text{SO}_4$, (Matthias and Remeika, 1956); ammonium fluoberyllate, $(\text{NH}_4)_2\text{BeF}_4$, (Pepinsky and Jona, 1957); triglycine sulfate (TGS), $(\text{CH}_2\text{NH}_2\text{COOH})_3\text{H}_2\text{SO}_4$, (Matthias, et al. 1956); triglycine fluoberyllate (TGF), $(\text{CH}_2\text{NH}_2\text{COOH})_3\cdot\text{H}_2\text{BeF}_4$, (Pepinsky, et al. 1957); and some other glycine compounds.

Of the many new compounds, one class deserves special attention. Solomon and Goldsmith (1956) found thiourea, $(\text{NH}_2)_2\text{CS}$, a material with a molecular crystal structure, to be ferroelectric. It is the only molecular crystal known to be ferroelectric.

In discussing the properties of ferroelectrics it is convenient to make a distinction between soft and hard ferroelectrics. Soft ferroelectrics are water-soluble, mechanically soft and have a low melting or decomposition temperature (RS, KH_2PO_4 , GASH, TGS, etc.). Hard ferroelectrics are usually grown at high temperatures from the melt or a flux; they are mechanically hard and not water-soluble (BaTiO_3 , KNbO_3 , $\text{Cd}_2\text{Nb}_2\text{O}_7$, etc.).

3.2. State of Knowledge

The phenomenology of ferroelectrics and antiferroelectrics¹ is quite well understood; the spontaneous polarization and the piezoelectric, elastic, and optical anomalies can be regarded as consequences of the anomalous dielectric properties. This applies to materials showing first- or second-order transitions at the Curie point.

The molecular mechanism, however, is still very little understood. Many properties have been investigated thoroughly, and many facts can be explained; but fundamental problems still remain which have to be understood. For convenience the problems will be lumped together into two main questions. First, what materials are ferroelectric or antiferroelectric and why? Second, how is the polarization reversed (that is, how are new domains nucleated and how do they grow)? To answer the first question

¹ In the antiferroelectric state, there exists a spontaneous polarization but the orientation of the dipoles on neighboring lines of ions is such that the net dipole moment of the array is zero. The simplest arrangement is one in which neighboring lines are polarized in antiparallel directions.

by Holden, Matthias, Merz, and Remeika (1953, 1956), however, turned out to be the major step in the discovery of new ferroelectrics. It became clear that many compounds with hydrogen bonds O-H-O and O-H-N are ferroelectric, especially when these bonds are linking groups like $(\text{SO}_4)^{--}$, $(\text{SeO}_4)^{--}$ etc. Matthias and Pepinsky and others have discovered many new ferroelectrics during the last few years; some of them are: alums (Pepinsky, et al., 1956); ammonium sulfate, $(\text{NH}_4)_2\text{SO}_4$, (Matthias and Remeika, 1956); ammonium fluoberyllate, $(\text{NH}_4)_2\text{BeF}_4$, (Pepinsky and Jona, 1957); triglycine sulfate (TGS), $(\text{CH}_2\text{NH}_2\text{COOH})_3\text{H}_2\text{SO}_4$, (Matthias, et al. 1956); triglycine fluoberyllate (TGF), $(\text{CH}_2\text{NH}_2\text{COOH})_3 \cdot \text{H}_2\text{BeF}_4$, (Pepinsky, et al. 1957); and some other glycine compounds.

Of the many new compounds, one class deserves special attention. Solomon and Goldsmith (1956) found thiourea, $(\text{NH}_2)_2\text{CS}$, a material with a molecular crystal structure, to be ferroelectric. It is the only molecular crystal known to be ferroelectric.

In discussing the properties of ferroelectrics it is convenient to make a distinction between soft and hard ferroelectrics. Soft ferroelectrics are water-soluble, mechanically soft and have a low melting or decomposition temperature (RS, KH_2PO_4 , GASH, TGS, etc.). Hard ferroelectrics are usually grown at high temperatures from the melt or a flux; they are mechanically hard and not water-soluble (BaTiO_3 , KNbO_3 , $\text{Cd}_2\text{Nb}_2\text{O}_7$, etc.).

3.2. State of Knowledge

The phenomenology of ferroelectrics and antiferroelectrics¹ is quite well understood; the spontaneous polarization and the piezoelectric, elastic, and optical anomalies can be regarded as consequences of the anomalous dielectric properties. This applies to materials showing first- or second-order transitions at the Curie point.

The molecular mechanism, however, is still very little understood. Many properties have been investigated thoroughly, and many facts can be explained; but fundamental problems still remain which have to be understood. For convenience the problems will be lumped together into two main questions. First, what materials are ferroelectric or antiferroelectric and why? Second, how is the polarization reversed (that is, how are new domains nucleated and how do they grow)? To answer the first question

¹ In the antiferroelectric state, there exists a spontaneous polarization but the orientation of the dipoles on neighboring lines of ions is such that the net dipole moment of the array is zero. The simplest arrangement is one in which neighboring lines are polarized in antiparallel directions.

we have to know more about the chemical bonds, the structure, and the internal field; to answer the second question we have to know more about surface layers and space charge in the crystal. In the following paragraphs some of these properties will be discussed.

3.2.1. Type of Bonds in Soft Ferroelectrics

Very careful X-ray work by a number of people, especially by Pepinsky and his coworkers, on KH_2PO_4 , RS, TGS, and other materials has shown that in practically all the soft ferroelectrics, ferroelectricity has a two-part cause: first, the dipole moment is formed by a shift of certain ions, primarily by the deformation of atomic groups like $(\text{SO}_4)^{--}$, $(\text{SeO}_4)^{--}$, $(\text{BeF}_4)^{--}$, $(\text{PO}_4)^{---}$, and $(\text{AsO}_4)^{---}$. In these groups the central ion (S, Se, Be, P, As, etc.) is shifted in the Curie region from the center into an eccentric position. Second, the dipoles thus created are aligned spontaneously by the action of the hydrogen bonds, making ferroelectricity possible. Coming from high temperatures, the hydrogen bonds at the Curie point go into an ordered state, thus causing the same distortion of all the atomic groups by a shift of the central ions in the same direction. This "trigger" action of the hydrogen bonds was first proposed by Slater (1941) for KH_2PO_4 and verified with X-rays by Frazer and Pepinsky (1953). It is important to note that in this case the hydrogen bond itself has no dipole moment in the ferroelectric direction (it is perpendicular to it). A number of other soft ferroelectrics are being investigated in several laboratories in order to verify this model.

An exception in the group of soft ferroelectrics is thiourea. It is a molecular crystal so that the molecule itself must have a dipole moment. Goldsmith and White (in press) have shown that the coupling between the dipoles causing ferroelectric properties takes place by Van der Waals forces.

3.2.2. Type of Bonds in Hard Ferroelectrics.

The best known of the hard ferroelectrics is BaTiO_3 , which crystallizes in the perovskite structure. In all perovskite and pyrochlore type ($\text{Cd}_2\text{Nb}_2\text{O}_7$) ferroelectrics, the key building block of the structure is the oxygen octahedron. Above the Curie point the octahedra are symmetrical. The dipole moment (and thus ferroelectricity) is caused in these materials by the deformation of the oxygen octahedra, which can be compared with the deformation of the $(\text{SO}_4)^{--}$, $(\text{PO}_4)^{---}$, etc., groups in the soft ferroelectrics. On the other hand, in the hard ferroelectrics the coupling be-

tween the dipoles which makes a spontaneous polarization possible is not caused by hydrogen bonds but by sharing of oxygen ions. Each octahedron is connected with six other octahedra by sharing one of the corners. This is a very different mechanism from the hydrogen bonds. The bonds within the oxygen octahedra are partially ionic and partially covalent. A purely ionic bond certainly would favor undistorted octahedra. This then suggests that going through the Curie point into the ferroelectric region, the directional (covalent) bond type becomes more powerful, tending to form angles which correspond to a distorted structure of lower symmetry (Megaw, 1954).

3.2.3. Structure and Structure Changes

In all ferroelectrics the structure symmetry is lowered by passing through the Curie point toward lower temperatures. The symmetry below the Curie point must be polar, whereas above it, it can be, but need not, be polar. The hard ferroelectrics are usually cubic above the Curie point, thus showing no polar axis and no piezo effect. The transition itself can be an order-disorder transition like the ordering of the hydrogen bonds in KH_2PO_4 , or it can be a displacive transition by shifting certain ions as in BaTiO_3 . Most of the soft ferroelectrics which have hydrogen bonds show a transition entropy of the order of magnitude of $1 \text{ cal-mole}^{-1}\text{-degree}^{-1}$ which is independent of their Curie point (Känzig, 1957). This rather large value indicates an order-disorder character of transition; RS seems to be an exception (Känzig, 1957). On the other hand, the hard ferroelectrics usually show a small transition entropy which is typical for displacive transitions (shift of central ion in oxygen octahedra).

Since the transition from the ferroelectric state to the paraelectric state is accompanied by a change in crystallographic structure, it has become a powerful method to look for ferroelectricity in crystals which are known to show a phase transition. There is a good chance of finding ferroelectricity or antiferroelectricity in insulators which show a phase transition.

3.2.4. Dipoles and Interaction of Dipoles

Practically all soft ferroelectrics with the exception of the alums have only one axis of polarization (up or down). The hard ferroelectrics, on the other hand, due to their cubic symmetry above the Curie point can polarize along several axes that are equivalent in the nonpolar state. In BaTiO_3 , for example, one finds domains at room temperature whose polarization is parallel

or antiparallel along any of the three crystallographic axes. This makes it easier to see the ferroelectric domains. The domain structure in BaTiO_3 has been investigated very carefully by Forsbergh (1949), Merz (1954), Hooton and Merz (1955), Little (1955), Miller (1959) and others. Zwicker and Scherrer (1944) and Mitsui and Furuichi (1952) have investigated the domain structure in KH_2PO_4 , Mitsui, Furuichi, and Marutake (1952) in RS, and Pearson and Feldmann (1959) in TGS.

An important result of all of these investigations is that the coupling of the electric dipoles in the forward direction is much larger than in the perpendicular direction. A consequence of this is that, in a number of cases, there is energetically not much difference between a ferroelectric and an antiferroelectric dipole arrangement. According to Cohen (1951), an antiferroelectric array in BaTiO_3 would be more stable than a ferroelectric array; Kinase (1955) and Takagi (1953) arrived at the opposite result. This certainly indicates that for both modifications the free energy is about the same. This statement is particularly true for the perovskites. Interestingly enough, Shirane and his coworkers (1951) were able to transform PbZrO_3 from an antiferroelectric into a ferroelectric state by applying a strong d.c. field. Similar results were obtained by Cross and Nicholson (1955) on NaNbO_3 . Another consequence of the strong coupling in the forward direction but weak coupling perpendicular to it is that the ferroelectric domains usually do not like to grow by sidewise motion of the walls. In other words, in contrast to ferromagnetic crystals, the polarization is reversed by nucleation of many new domains instead of the growth of very few existing domains by sidewise motion of the walls.

The nucleation of new domains depends on the number of existing domains. In many ferroelectrics, new nuclei are primarily formed next to existing ones; that is, there is a strong domain-nucleus interaction. Fatuzzo and Merz (in press) have shown that this interaction is usually high if the spontaneous polarization is high.

The dielectric behavior of ferroelectrics is quite well understood. Below the Curie point there is not a linear relation between applied electric field and resulting electric polarization, and one observes a hysteresis loop. With increasing temperature the dipole moment becomes smaller. At the Curie point the polarization drops to zero, either because the dipole moment itself becomes zero, or because the existing dipoles lose their ordered state, or both. The transition can be of first- or second-order, and it can be a

displacive or an order-disorder transition. At the Curie point the dielectric constant goes through a very sharp maximum, which is simply a manifestation of the fact that certain ions can be shifted very easily by an applied electric field.

For the same reason, all ferroelectric materials which are not cubic above the Curie point show a very large piezoeffect at that point. Because the change of crystal structure changes or creates electrical dipoles and vice versa, a change of polarization at the Curie point is coupled with a deformation. An interesting consequence of this is that a completely clamped crystal cannot become ferroelectric when it is cooled below the Curie point. This has been shown by Baumgartner (1951) for KH_2PO_4 .

A ferroelectric crystal is pyroelectric in the ferroelectric range. The pyroelectric effect is proportional to the temperature dependence of the spontaneous polarization. Chynoweth (1956) has used this effect to measure the spontaneous polarization of ferroelectrics. This method is very useful when a very high coercive force prevents switching of the polarization.

Quite often the ferroelectric hysteresis loop of certain materials is strongly biased or distorted. Holden, Merz, Remeika, and Matthias (1956) have shown that the bias in GASH seems to be connected with the growth of the crystal. Chynoweth (1959) has shown that the bias in TGS can be obtained by electron bombardment, whereby the direction of bias is connected with the direction of the polarization. Jurin (1957) and Fatuzzo (private communication) have shown that bias and distortion can be caused by impurities.

3.3. Problems Needing Further Research—Some Proposals

In this chapter the problems which seem to be the key problems in the field of ferroelectrics will be discussed. Listed are some specific problems which need further research and some specific experiments and ideas which might lead to a better understanding of ferroelectricity. The key problems can be summarized in the following two questions. First, what materials are ferroelectric and why? Second, how are domains formed and how do they grow?

3.3.1. What Materials Are Ferroelectric and Why?

During the last few years tremendous progress has been made in discovering new ferroelectrics and even in predicting their occurrence. Since the discovery of GASH in 1955, Matthias and Pepinsky and others have discovered more than a dozen new types

of ferroelectrics. However, all of these new materials are soft ferroelectrics, that is consisting of deformable molecular groups like $(\text{SO}_4)^{--}$, $(\text{SeO}_4)^{--}$, $(\text{SeO}_3)^{--}$, etc., and hydrogen bonds like O-H-O and O-H-N. The electrical properties and structures of most of these new materials are now being investigated.

In the future, however, more emphasis should be put on finding new hard ferroelectrics. Few are known, and it is very likely that some hard ferroelectrics with a simple structure do exist. Due to their mechanical strength and high electrical polarization, the hard ferroelectrics might be much more important for technical applications (BaTiO_3 is still the fastest ferroelectric switch). In both the perovskite and the pyrochlore type of hard ferroelectrics, the oxygen octahedra are the main feature of the structure; their deformation gives the dipole moment, and the sharing of the oxygen ions gives the coupling. An important question is whether other atomic arrangements besides octahedra can lead to ferroelectricity in hard ferroelectrics.

Thiourea, as a molecular crystal, is a very special compound representing a completely different type of mechanism for ferroelectricity. Since not much is known about this type of dielectric coupling it would be desirable to discover some other ferroelectrics with a molecular crystal structure.

Finding new ferroelectrics is important in order to get more information about ferroelectricity. It is at least as important, however, to learn more about the internal field throughout the crystal and about the vibrational spectrum of the ions in the crystal and their temperature dependence. At the Curie point one expects to find a marked change in the infrared absorption. But La Lau (1947), Lord and Merrifield (1953), Murphy, Weiner, and Oberly (1954), and Chapelle (1949) could not find any significant change in the infrared absorption with KH_2PO_4 , and Mara (1954) and Last (1957) did not find any of BaTiO_3 . Chapelle (1949) found only a small change in the Raman spectrum in RS. It is quite possible that all the measurements were made at too short wavelengths (1μ – 30μ). On theoretical considerations, Anderson (1958) predicted a very pronounced change in the infrared absorption spectrum for BaTiO_3 in the millimeter wavelength range. No measurements have been made in this region, and it would be desirable if these experiments were performed.

It is difficult to get theoretical information about the internal field. The balance of forces which determine the symmetry of the crystal, that is the dipoles and their interaction, is very delicate. As long as normal dielectrics are not better understood

it will be almost impossible to understand ferroelectrics or antiferroelectrics. On the other hand, the first-order anomalies which can be observed experimentally in ferroelectrics might well be very important, not only for the theory of ferroelectrics but also for the theory of normal dielectrics.

Some experiments have been performed in order to get more information about the internal fields. A number of investigators have been looking for Stark-splitting without success. Paramagnetic resonance work has been done on GASH (Bogle, et al., 1957) with Cr in the Al positions and on BaTiO₃ (Hornig, et al., 1959) with Fe in the Ti positions. At the Curie point a splitting of lines due to the change of symmetry could be observed in agreement with X-ray work. Proton magnetic resonance work has been done on GASH by Spence and Muller (1957) and by McCall (1957), on thiourea by Emsley and Smith (1959) and Goldsmith and White, (private communication) and on KH₂PO₄ by Newman (1950). Some interesting quadrupole resonance experiments of Nb in KNbO₃ have been performed by Cotts and Knight (1954). They found some very pronounced changes of the field gradient ∇E at the transition points. At the Curie point, ∇E disappears because the crystal becomes cubic; in agreement with X-ray data it appears that, above the Curie point, the Nb is centrally located in the oxygen octahedra. Furthermore, Cotts and Knight found that the gradient, ∇E , in the polarized state is too large to be explained by a purely ionic model. The bonding has to be partially covalent in agreement with the proposal of Megaw (1954) and others. Burns (private communication) has observed interesting effects in the quadrupole resonance in GASH and other ferroelectrics. These latter results clearly demonstrate how useful resonance work can be; much more should be done along these lines.

3.3.2. Domain Formation and Domain Wall Motion

The domain formation (nucleation) and the domain wall motion has been studied on BaTiO₃ by Merz (1954, 1956), Little (1955), Wieder (1955), Landauer et al. (1956), Chynoweth (1958), Miller (1958), and others; on GASH by Wieder (1957) and Prutton (1957), on RS by Wieder (1958), and on TGS by Pulvari and Kuebler (1952) and Fatuzzo and Merz (in press). It is generally agreed that the switching of polarization in ferroelectrics is a two-step process; first, nucleation of new domains at the surface of the sample, and second, growth of the domains by wall motion. However, very many questions are still unanswered.

In an insulator the space charge always plays an important

role. In a ferroelectric it probably influences the switching greatly. It is even conceivable that space charge might be carried into the crystal on the domain walls. The very pronounced bias in the hysteresis loop of GASH and colmanite must be due to space charge which acts like an internal battery. This was shown by the experiments mentioned above by Jurin (1957), Chynoweth (1959), and Fatuzzo. Chynoweth's observation that the bias induced by electron bombardment depends on the polarization offers a powerful approach to the study of the role of space charge.

In this connection it has to be mentioned that the problem of fatigue might also be connected with space charge. A number of ferroelectric materials, BaTiO_3 in particular, under certain pulsing conditions show very pronounced fatigue. This manifests itself in the smaller signals obtained when the polarization is reversed; that is, not all the domains can be switched. The question is whether in this condition it is impossible to nucleate enough domains because of a surface charge layer that is formed, or whether the domain walls cannot be moved through the crystal because of space charge inside the crystal. In BaTiO_3 there is some evidence (Anderson, et al., 1955) that the latter occurs. Due to the high currents which flow through the crystal when switched, there might occur some electrolysis which by reduction transforms some of the Ti^{4+} ions into Ti^{3+} ions plus free charges, giving rise to space charge. The whole field of surface chemistry and electrolysis in ferroelectrics requires much more careful study.

A more detailed study of the size and shape of the domain nuclei, of the wall thickness, and of the wall energy would be very desirable too. Such studies would allow a more detailed calculation of the activation field for nucleation and of the mobility for wall motion.

The high frequency properties of the dielectric constant are also of great interest. The main question is at what frequency the dielectric constant drops to low values. Yager (private communication) finds the dielectric constant of KH_2PO_4 still high at 2.5×10^{10} cps and room temperature. Measurements of dielectric relaxation in BaTiO_3 and other perovskites have shown that the relaxation frequencies in the tetragonal polarized phase seem to be between 10^7 and 3×10^9 cps. However, this relaxation is probably due to domain wall oscillations as calculated by Kittel (1951). This theory is substantiated by the experiments of Benedict and Durand

(1958) who measured the dielectric constant on single crystals which were poled and did not have any domain walls. At 24 KMc/sec they were still far away from any ionic resonance. This fact, together with the experiments and predictions on infrared absorption (reported above), make it imperative to look at the absorption properties of ferroelectrics in the millimeter wave range. There must be a relaxation! The high frequency measurements have also shown that all ferroelectrics become very lossy at high frequencies. This has been a serious limitation of high frequency applications of ferroelectrics. Although it is conceivable that this is due to domain wall oscillations (Kittel, 1951), much more experimentation along these lines is needed.

3.4. References

- Anderson, J. R., Brady, G. W., Merz, W. J., and Remeika, J. P., 1955, *J. Appl. Phys.*, **26**, 1387.
- Anderson, P. W., 1958, *International Conference on Dielectrics*, Moscow.
- Baumgartner, H., 1951, *Helv. Phys. Acta*, **24**, 326.
- Benedict, T. S., and Durand, J. L., 1958, *Phys. Rev.*, **109**, 1091.
- Bogle, G. S., Gabriel, J. R., and Bottomley, G. A., 1957, *Trans. Farad. Soc.*, **53**, 1058.
- Burns, J., private communication.
- Busch, G., and Scherrer, P., 1935, *Helv. Phys. Acta*, **20**, 273.
- Chapelle, J., 1949, *Thesis*, University of Paris.
- Chynoweth, A. G., 1956, *J. Appl. Phys.*, **27**, 78.
- , 1956, *Phys. Rev.*, **102**, 705, 1021.
- , 1958, *Phys. Rev.*, **110**, 1316.
- , 1959, *Phys. Rev.*, **113**, 159.
- Cohen, M. H., 1951, *Phys. Rev.*, **84**, 369.
- Cook, W. R., and Joffe, H., 1952, *Phys. Rev.*, **88**, 1426.
- Cotts, R. M., and Knight, W. D., 1954, *Phys. Rev.*, **96**, 1285.
- Cross, L. E., and Nicholson, B. J., 1955, *Phil. Mag.*, **46**, 453.
- Drougard, M. E., and Landauer, R., private communication.
- Emsley, Smith, 1959, *Colloque Ampere*, London, April.
- Fatuzzo, E., and Merz, W., in press.
- , private communication.
- Forsbergh, P. W., 1946, *Phys. Rev.*, **76**, 1187.
- Frazer, B. C., 1953, *Phys. Rev.*, **91**, 212.
- , and Pepinsky, R., 1953, *Acta Cryst.* **6**, 273.
- Goldsmith, G. J., and White, J. G., in press.
- , and ———, private communication.
- Hablützel, J., 1939, *Helv. Phys. Acta*, **12**, 489.
- Hippel, von A., 1950, *Rev. Mod. Phys.*, **22**, 221.
- Holden, A. N., Matthias, B. T., Merz, W., and Remeika, J. P., 1953, *Phys. Rev.*, **93**, 546.
- , Merz, W., Remeika, J. P., and Matthias, B. T., 1956, *Phys. Rev.* **101**, 962.

- Hooten, J. A., and Merz, W., 1955, *Phys. Rev.*, **98**, 409.
Hornig, A. W., Rempel, R. C., and Weaver, H. E., 1959, *J. Phys. Chem. Solids*, **10**.
Jurin, V. A., 1957, *Bull. Acad. Science USSR*, **21**, 829.
Känzig, W., 1955, *Phys. Rev.*, **98**, 549.
———, 1957, *Solid State Physics*, Vol. 4 (Academic Press, New York).
Kinase, W., 1955, *Prog. Theoret. Phys.*, Japan, **13**, 529.
Kittel, C., 1951, *Phys. Rev.*, **83**, 458.
La Lau, C., 1947, *Thesis*, University of Amsterdam.
Landauer, R., Young, D. R., and Drougard, M. E., 1956, *J. Appl. Phys.*, **27**, 752.
Last, J. T., 1957, *Phys. Rev.*, **105**, 1740.
Little, E. A., 1955, *Phys. Rev.*, **98**, 978.
Lord, R. C., and Merrifield, R. E., 1953, *J. Chem. Phys.*, **21**, 166.
Matthias, B. T., Merz, W., and Scherrer, P., 1947, *Helv. Phys. Acta*, **20**, 273.
———, and Remeika, J. P., 1951, *Phys. Rev.*, **82**, 727.
———, and ———, 1956, *Phys. Rev.*, **103**, 262.
———, 1952, *Phys. Rev.*, **85**, 723.
———, Miller, C., and Remeika, J. P., 1956, *Phys. Rev.*, **104**, 849.
Mara, R. T. et al., 1954, *Phys. Rev.*, **96**, 801.
Marutake, M., 1952, *J. Phys. Soc.*, Japan, **7**, 25.
McCall, D. W., 1957, *J. Chem. Phys.*, **26**, 706.
Megaw, H., 1954, *Acta Cryst.*, **7**, 187.
Merz, W., 1954, *Phys. Rev.*, **95**, 690.
———, 1956, *J. Appl. Phys.*, **27**, 938.
Miller, R. C., 1958, *Phys. Rev.*, **111**, 736.
———, and Savage, A., 1959, *Phys. Rev. Letters*, **2**, 294.
Mitsui, T., and Furuichi, J., 1952, *Rev.*, **90**, 193.
Murphy, G. M., Weiner, G., and Oberly, J. J., 1954, *J. Chem. Phys.*, **22**, 1322.
Newman, R., 1950, *J. Chem. Phys.*, **18**, 669.
Pearson, G. L., and Feldman, W. L., 1959, *J. Phys. Chem. Solids*, **9**, 21.
Pepinsky, R., Jona, F., and Shirane, G., 1956, *Phys. Rev.*, **102**, 1181.
———, and ———, 1957, *Phys. Rev.*, **105**, 344.
———, et al., 1957, *Phys. Rev.*, **107**, 1255.
Pruett, M., 1957, *Proc. Roy. Soc.*, **70B**, 1064.
Pulvari, C. F., and Kuebler, W., 1958, *J. Appl. Phys.*, **29**, 1742.
Shirane, G., et al., 1951, *J. Phys. Soc.*, Japan, **6**, 265.
———, Sawaguchi, E., and Tagaki, Y., 1951, *J. Phys. Soc.*, Japan, **6**, 333.
———, Sawaguchi, E., and Tagaki, Y., 1951, *Phys. Rev.*, **84**, 476.
Slater, J. C., 1941, *J. Chem. Phys.*, **9**, 16.
Solomon, A., and Goldsmith, G. J., 1956, *Phys. Rev.*, **104**, 1191.
Spence, R. D., and Muller, J., 1947, *J. Chem. Phys.*, **26**, 706.
Tagaki, J., 1953, *Proc. Int. Conf. Theor. Phys.*, Kyoto and Tokyo, p. 284.
Valasek, J., 1921, *Phys. Rev.*, **17**, 475.
Wieder, H. H., 1955, *Phys. Rev.*, **99**, 1161.
———, 1957, *Proc. Inst. Radio Eng.*, **45**, 1094.
———, 1958, *Phys. Rev.*, **110**, 29.
Wul, B., 1946, *J. Phys.*, USSR, **10**, 95.
Yager, W. A., private communication.
Zwicker, B., and Scherrer, P., 1944, *Helv. Phys. Acta*, **17**, 346.

4. THERMOELECTRICITY

WAYNE W. SCANLON

*U. S. Naval Ordnance Laboratory
Silver Spring, Maryland*

4.1. Introduction

The transport of electrical and thermal energy in a material results in a number of interrelated properties called thermoelectric effects which may be attributed to the mutual interference of heat flow and electric current flow. There are three principal thermoelectric effects which are defined as follows:

1. Seebeck Effect. If two different conductors are joined together at both ends and the two junctions kept at different temperatures, an electromotive force is set up which is proportional to the temperature difference for small temperature differences. The thermoelectromotive force per degree is called the thermoelectric power, S , or Seebeck coefficient.

2. Thomson Effect. When an electric current, I , passes between two points of a homogeneous conductor, with a temperature difference ΔT existing between these points, an amount of heat $\tau I \Delta T$ is emitted or absorbed in addition to the Joule heat. The parameter τ is called the Thomson coefficient.

3. Peltier Effect. If two conductors are joined together and kept at a constant temperature while a current, I , passes through the junction, heat is generated or absorbed at the junction in addition to the Joule heat. The Peltier coefficient Π_{12} is defined as the amount of heat, per unit of electric current, emitted or absorbed per second at the isothermal junction of two materials.

4.1.1. The Kelvin Relations

The three thermoelectric effects are related through the Kelvin relations

$$S = \frac{\pi}{T} \quad (4.1)$$

and

$$\frac{dS}{dT} = \frac{(\tau_1 - \tau_2)}{T} \quad (4.2)$$

Although Kelvin pointed out that the method of derivation of the above equations is not completely justifiable, experiments confirm their validity. The exact status of the Kelvin thermodynamic relations among the various thermoelectric phenomena in con-

ductors has long been a matter for discussion. Onsager (1931, 1945) developed a set of reciprocal relations for treating the symmetry in the mutual interference of two or more irreversible processes occurring simultaneously in a system. He showed that the whole class of results obtained by Kelvin follows from the reciprocal relations. H. B. Callen (1948) analyzed the thermoelectric effects by means of the Onsager reciprocal relations and obtained a simple and intuitive interpretation of the thermoelectric effects and a valid derivation of the Kelvin relations.

More recently Sondheimer (1956) reviewed critically the application of the symmetry relations to thermoelectric phenomena. He pointed out that some of the recent theoretical work on thermoelectric effects in semiconductors has been formulated in a way which violates the Kelvin relations. The proof of the Kelvin relations given by Sondheimer does not depend upon the properties of any particular model. It is necessary in practice, however, to introduce many simplifying assumptions in calculating the various thermoelectric effects. In order to avoid conflicts with the Kelvin relations it is important to treat the mutual interference of electrons and lattice vibrations in a fully symmetric fashion. The consequences of this requirement should be explored more fully.

4.1.2. Absolute Thermoelectric Power

Thermoelectric power generally refers to two conductors since a pair is needed to complete a circuit. However, the thermoelectric effect is characteristic of the bulk material, and it is possible to define an absolute thermoelectric power for a material. If this information is known for one material then by comparing it with other materials their absolute thermoelectric powers can be obtained. Absolute thermoelectric power data have been obtained on lead (Borelius, et al., 1932) from room temperature down to about 20°K, based upon direct measurement of the Thomson heat. Below 20°K absolute thermoelectric data were obtained by using a superconductor as a reference material (Jan, et al., 1958; Christian, et al., 1958).

4.2. Thermoelectric Effects in Metals

Formulas for the absolute thermoelectric power have been derived based upon familiar forms of the electron theory of conduction. In the case of metals the analysis yields the following expression for the absolute thermoelectric power (Mott and Jones, 1936; Wilson, 1954):

$$S = \frac{\pi^2 k^2 T}{3e} \left\{ \frac{\delta \log \sigma(E)}{\delta E} \right\} \quad E = E_0, \quad (4.3)$$

where E_0 is the Fermi energy at $T = 0^\circ\text{K}$ and σ is the conductivity. The conductivity depends upon the scattering mechanisms, the density of carriers and the energy band structure.¹

At high temperatures ($T > \theta$, where θ is the Debye temperature) the behavior is classical in nature, a relaxation time always exists, and scattering by impurities is usually small or negligible. In this region fairly definite theoretical conclusions may be derived regarding the thermoelectric power. Wilson (1954) shows that Eq. (4.3) may be written as

$$S = \frac{\pi^2 k^2 T}{3eE_0} x \quad (4.4)$$

where x can be evaluated for various scattering mechanisms and constant energy surfaces. For the case of perfectly free electrons and spherical energy surfaces Wilson finds that $x = 3$.

Experimental studies on monovalent metals like Cu, Na, K, and others which are generally considered to be close approximations to an ideal free electron model show that x varies widely from one metal to another (MacDonald and Roy, 1953). The values obtained suggest that there is considerable deviation from the idealized free electron model in the alkali and noble metals.

At low temperatures the theory becomes simplified since the mean free path is determined by impurities and is independent of E and T , and we have

$$S = \frac{\pi^2 k^2 T}{3eE_0} \quad (4.5)$$

Experiments on the alkali and noble metals at low temperatures (ranging from 2° to 20°K) show large discrepancies from the theoretical behavior both in sign and by as much as a factor of 100 in magnitude (MacDonald, et al., 1953, 1958). Even in sodium and lithium, which one might hope to be nearest to the idealized model, the deviations from theory are serious. Furthermore, the results for the various metals differ greatly from one another in their dependence upon temperature, and the discrepancy with theory becomes progressively greater in metals as their atomic weight increases.

One possible source for the additional thermoelectric power found experimentally in metals lies in the phonon-drag effect

¹ Phonon drag is not included in the analysis leading to Eq. (4.3). The resolution of the paradox that the energy level width is often greater than kT is given by Lewis (1958).

(Gurevich, 1945, 1946). However, there are conflicting results for some metals. In copper, for example, impurities tend to increase, rather than decrease, the thermoelectric power at low temperatures and in other metals such as cesium and rubidium the sign of the effect may not be accounted for by the phonon drag mechanism.

Sign reversals of the thermoelectric power at low temperatures could result from electron-phonon collisions of the so-called Umklapp-type where momentum reversal may occur. These processes are generally thought to be negligible. However, recent work indicates that they have been greatly underestimated (Bailyn, 1958), and Umklapp processes are estimated to contribute substantially to the thermoelectric power. The analysis of the contribution of phonon drag and Umklapp processes in metals is complex and has not yet been worked out quantitatively.

4.3. Thermoelectric Effects in Semiconductors and Insulators

Thermoelectric effects in semiconductors have been studied with somewhat less intensity than other properties of semiconductors. The thermal effects are more complex in semiconductors than in metals due to the possibility of charge carriers of different signs, statistical behavior ranging from classical to Fermi statistics, the presence of a variety of scattering mechanisms, and sometimes more complicated transport mechanisms like charge jumping from one local site to another.

Theories of thermoelectric effects in semiconductors have been developed from two different approaches. In one the tendency of charge carriers to diffuse from hot to cold regions in a conductor is analyzed by solving the Boltzman transport equations for a variety of assumptions regarding scattering processes, effective mass, statistics, and for different models of constant energy surfaces. An excellent review of this treatment of the problem is given by Johnson (1956). Agreement with experiment is generally good for the semiconductors Ge and Si, particularly at high temperatures.

Another method for deriving expressions for thermoelectric effects is due to C. Wagner who applied the Kelvin relation between thermoelectric power and the Peltier heat, Eq. (4.1). This method is discussed by Herring (1954, 1958) and developed for an n-type semiconductor of low carrier density. An expression for the thermoelectric power used in the analysis is,

$$S = \mp \frac{\epsilon_b - \epsilon_F}{eT} + \alpha \frac{k}{e} \quad (4.6)$$

in which the $+$ sign is for holes, the $-$ sign is for electrons, ϵ_c is the energy of the conduction or valence band edge, respectively, and α is the fraction by which the average energy of carriers is above the band edge. This expression is valid for any band structure. To apply Eq. (4.6) to a given semiconductor it is necessary to locate the Fermi energy. This requires assumptions about the effective mass and band degeneracies which are accurate for materials like germanium and silicon. Reasonable estimates can be used for other materials in the absence of accurate data for approximate solutions. The value of α for any band where energy is a homogeneous quadratic function of crystal momentum is 2 for lattice scattering and slightly greater or less for other scattering mechanisms.

At low temperatures electrons and holes transport insufficient energy to account for all the Peltier heat. At these temperatures the transport of energy by the lattice vibrations must be considered. The mutual interference of electron and phonon currents results in a drag of one on the other, the Gurevich effect, which contributes to the thermoelectric power and Peltier heat. The contribution of the phonon drag to thermoelectric power is discussed by Frederikse (1953) and Geballe and Hull (1954). The drag effect upsets the phonon distribution, and since collision processes which attempt to restore the phonons to equilibrium are less effective at low temperatures, the phonon drag contribution to the thermoelectric power becomes increasingly effective at low temperatures.² Experiments tend to give a weaker temperature dependence than the $T^{-7/2}$ relation. The magnitude of the phonon drag contribution is not readily calculated because of some uncertainty in scattering constants; however, it may be fairly large in germanium even at room temperature (Herring, 1958). The phonon drag effect upon the thermoelectric power decreases with increasing carrier concentration.

At sufficiently low temperatures (about 10°K) additional scattering mechanisms, such as boundary scattering, become important which cause a decrease in the thermoelectric power.

At carrier concentrations just below that for degeneracy, impurity band conduction may exist. The sign of the thermoelectric power may not be the same in the impurity band as for free carriers. In some Si samples, for example, Geballe and Hull (1955) observed a reversal of sign of S in the range from 15°

² A $T^{-7/2}$ dependence has been derived for a model which may not, however, be adequate.

to 20°K which they attributed to a transition from free carrier conduction to impurity band conduction.

The agreement of the theory of thermoelectric power with experiment is generally good for the best known semiconductors, germanium and silicon, since many of the detailed features of scattering, effective masses, and band structure are known from independent sources. The interpretation of thermoelectric data involves a great number of different properties of a given material, and for many other semiconductors little is known of some of these properties, although one can make reasonable estimates in some cases.

Thermoelectric effects in an insulator, diamond, have been studied (Goldsmid, et al., 1959). The results show that even at room temperature phonon drag effects are quite large (about 2.5 mv/°C) and the temperature dependence is nearly $T^{-7/2}$.

In the case of semiconductors or insulators in which carrier mobility is extremely small new concepts of conduction must be developed. One of these conduction mechanisms, applicable to a group of mixed valency compounds, is a jumping of charge from site to site in the lattice. With this type of conduction, the use of the energy band concept is questionable, and perhaps a localized model would be preferable. In one group of such materials, lithium-substituted transition metal oxides, the temperature variation of electrical conductivity is explained in terms of an activation associated with the mobility rather than with the number of charge carriers. In such a model the thermoelectric power based upon entropy of mixing is (Zener, 1959)

$$S = \frac{k}{e} \ln \left(\frac{1 - 2c}{c} \right) \quad (4.7)$$

where c is the atomic concentration of the impurity. This formula is confirmed on a number of doped transition metal oxides. One expects, however, an additional transport energy of the type familiar in the theory of the thermoelectric power of ionic crystals.

Other low-mobility semiconductors are In_2Se_3 , In_2Te_3 , Ga_2Se_3 , and Ga_2Te_3 with vacant metal ion sites.

A study of these low-mobility semiconductors has been somewhat neglected in favor of high-mobility materials. A continued effort to understand the physics of these semiconductors is of fundamental importance.

The greatest advances in understanding materials will probably continue to be based upon studies of single crystalline specimens of materials. In the transition metal oxides the problem of growing

single crystals is particularly severe, and the experimental studies reported in the literature are made upon polycrystalline sintered specimens. A strong effort should be directed toward obtaining single crystals in order to elucidate the transport properties of these low-mobility materials. An accompanying basic research program on low-mobility materials which are somewhat easier to prepare, such as the tellurides and selenides of indium and gallium, would seem appropriate.

4.4. Thermoelectric Conversion

Practical application of thermoelectric effects in the generation of electric power and in refrigeration have been receiving a great deal of attention in recent years. While it is the intention of this survey to emphasize fundamental aspects of thermoelectricity, it nevertheless seems appropriate to comment generally on some of the practical applications, since they often serve as a stimulus for pursuing basic studies.

The generation of electrical power by the use of the Seebeck effect is one of the oldest known generation methods. Ohm, for example, used a thermoelectric battery operating between 0° and 100°C in the experimental establishment of his law. A number of types of thermoelectric batteries were commercially available over a hundred years ago. One of these, for example, which used a zinc-antimony alloy thermocouple furnished about 3 volts with an internal resistance of 1.5 ohms and cost about \$15 or \$20 (Ganot, 1854).

With the revival of interest in this method for generating power we now have expressions for the efficiency of generators in terms of fundamental material properties. An excellent survey of this analysis is given by Ioffe (1956). An interesting result of the analysis is that the efficiency depends upon the values of three interrelated material parameters S , k , and p contained in a "figure of merit" z

$$z = \frac{S^2}{kp} \quad (4.8)$$

where k is the thermal conductivity and p is the electrical resistivity. Using appropriate expressions for S , k , and p for a simple semiconductor, one finds that an optimum efficiency in a semiconductor occurs for a thermoelectric power of about $2k/e$ or about $172\mu\text{V}/^\circ\text{C}$, which is near the onset of degeneracy.

In experiments the efficiency is substantially less than the theoretical value. This may be due in part to simplifications made in

the theory and in part to technological problems. The appropriateness of the simplified theory for materials which are degenerate or in which conduction occurs by a jump mechanism should be the subject of further study.

An understanding of transport phenomena in degenerate semiconductors and in semiconductors which have an activated mobility would help guide the selection of suitable materials for use in thermoelectric converters. Much remains to be done in this area of research. The knowledge available at present on materials suitable for thermoelectric batteries has led to thermoelectric-nuclear power supplies which have over ten times higher power-to-weight ratios than the best chemical batteries. With further knowledge the future looks bright for thermoelectric devices.

4.5. Other Thermoelectric Effects

Thermoelectric effects can be modified by a magnetic field. The magnitude of the thermoelectric power is changed, and an emf appears perpendicular to the thermal gradient and the magnetic field. The latter is called the Nernst effect and the former the magneto-Seebeck effect. From the dependence of these effects upon the orientation of the magnetic field and thermal gradient much useful information can be obtained on the constants relating to the anisotropy of electron-phonon interaction and the magnitude and frequency dependence of relaxation times of phonons.

The Nernst effect is due entirely to differences in the behavior of different groups of carriers and therefore is a very sensitive method for studying such things as scattering mechanisms and mass anisotropy. The theory of these effects is given by Herring (1958) and others. The analysis of the Nernst effect is complicated since all the details of electron and phonon conduction as well as their interactions are involved. The analysis and application to germanium has been made (Herring et al., 1958).

The magnetic field alters the thermoelectric power, principally through its effect on the phonon part of the thermoelectric power. The change in S may be written as the sum of two parts.

$$\Delta S = \Delta S_e + \Delta S_p$$

Theory predicts that ΔS_e cannot be much greater than k/e , whereas ΔS_p can be of the order of S_p .

In a magnetic field the thermoelectric power and Peltier coefficient are tensors. In n-type germanium there are only two principal components, one along the axis (\parallel) and one perpendicular to the axis (\perp). Herring (1958) discusses the significance of

the Peltier components, Π_{\perp} and $\Pi_{||}$ and their relation to the deformation potentials and the laws of phonon-phonon scattering.

The principal conclusions of the experimental studies of magneto-thermoelectric effects are that the theory is generally confirmed in n-type germanium, and the ratio $\Pi_{||}/\Pi_{\perp} \simeq m_{||}^* \tau_{||}/m_{\perp}^* \tau_{\perp} \simeq 9.6$ at 80°K. The average of $\Pi_{||}$ and Π_{\perp} decreases slowly with increasing energy which suggests that either the relaxation times of the transverse modes are at least as long as those of the longitudinal modes or that the longitudinal modes do not approximate the ideal behavior $\tau_p \approx \nu^{-2}$.

4.6. Conclusions

Although the experimental investigations of thermoelectric effects present no outstanding barriers, the theoretical interpretation of the phenomena is in general complicated, and the present theory is often in disagreement with the experiments. Quantitative comparisons between theory and experiment for many materials are not easily made due in part to the inadequacy of knowledge of many material properties including such things as scattering processes, band structure, and density of carriers. Even in the case of monatomic metals such as Cu and Na, which are generally thought to be close approximations to the ideal free electron model of a metal, the theory of thermoelectric effects sometimes shows wide divergencies from the experimental behavior both at high and low temperatures. For more complicated metals and increasing atomic weights, the discrepancy between theory and experiment becomes progressively greater.

In the case of high-mobility semiconductors the theory and experiment tend to be in reasonable agreement at high temperatures. At low temperatures in the simplest case of dilute n-type germanium, the theory agrees well with experiments. Quantitative agreement is still not generally found, due possibly to approximations made in the theory. The results of the low temperature studies have, however, contributed greatly to our understanding of electron-phonon and phonon-phonon interactions in solids. Further studies on thermoelectromagnetic phenomena may yield information on phonon-electron scattering anisotropy.

The theory of thermoelectric effects in low-mobility semiconductors, where conduction depends upon a charge-jumping mechanism, has been somewhat neglected, and a sound fundamental program of theoretical and experimental studies, preferably on good single crystals, is most important.

4.7. References

- Bailyn, M., 1958, *Phys. Rev.*, **112**, 1587.
 Borelius, G., Keesom, W. H., Johansson, C. H., and Linde, J. O., 1932, *Proc. Acad., Amsterdam*, **31**, (1), 25.
 Callen, H. B., 1948, *Phys. Rev.*, **73**, 1344.
 Christian, J. W., Jan, J. P., Pearson, W. B., and Templeton, I. M., 1958, *Proc. Roy. Soc.*, **A245**, 213.
 Frederikse, H., 1953, *Phys. Rev.*, **92**, 248.
 Ganot, 1854, *Elementary Treatise on Physics*, William Wood and Co., New York (Translated by E. Atkinson).
 Geballe, T. H., and Hull, G. W., 1954, *Phys. Rev.*, **94**, 1134.
 ———, and ———, 1955, *Phys. Rev.*, **98**, 940.
 Goldsmid, H. J., Jenne, C. C., and Wright, D. A., 1959, *Proc. Roy. Soc.*, **73**, 393.
 Gurevich, L., 1945, *J. Phys.*, Moscow, **9**, 477.
 ———, 1946, *J. Phys.*, Moscow, **10**, 87.
 Herring, C., 1954, *Phys. Rev.*, **96**, 1163.
 ———, 1958, *Semiconductors and Phosphors*, Friedr. Vieweg und Sohn, Braunschweig, 184.
 ———, Geballe, T. H., and Kunzler, J. E., 1958, *Phys. Rev.*, **111**, 36.
 Ioffe, A. F., 1956, *Semiconductor Thermoelements and Thermoelectric Cooling*, Infosearch, London.
 Jan, J. P., Pearson, W. B., and Templeton, I. M., 1958, *Canad. J. Physics*, **36**, 627.
 Johnson, V. A., 1956, *Progress in Semiconductors* (Ed. Gibson, Aigrain, Burgess), Haywood and Co., London.
 MacDonald, D. K. C., and Roy, S. K., 1953, *Phil. Mag.*, **44**, 1364.
 ———, and Pearson, W. B., 1953, *Proc. Roy. Soc.*, **A219**, 373.
 ———, Pearson, W. B., and Templeton, I. M., 1958, *Proc. Roy. Soc.*, **A248**, 107.
 Mott, N. F., and Jones, H., 1936, *Theory of the Properties of Metals and Alloys*, Oxford.
 Onsager, L., 1931, *Phys. Rev.*, **37**, 405; 1931, *Phys. Rev.*, **38**, 2265.
 ———, 1945, *Ann. N.Y. Acad. Sci.*, **XLVI**, 241.
 Sondheimer, E. H., 1956, *Proc. Roy. Soc.*, **A234**, 391.
 Wilson, A. H., 1954, *Theory of Metals*, Cambridge.
 Zener, C., 1959, *J. Phys. Chem. Solids*, **8**, 26.

5. THERMAL CONDUCTIVITY

P. G. KLEMENS

Westinghouse Research Laboratories
Pittsburgh, Pa.

5.1. Introduction

Heat conduction in solids takes place mainly by means of lattice waves and free electrons, but in principle any mode of excitation which contributes to the specific heat also contributes to the thermal conductivity, so that the latter is composed of contributions from various excitations (or carriers)

$$\kappa = \sum_i \kappa_i = \sum_i \alpha_i S_i v_i l_i \quad (5.1)$$

where α_i is a constant of order unity, S_i is the specific heat per unit volume due to carriers of type i , v_i their group velocity, and l_i an effective mean free path. However, interactions between carriers i and j will tend to reduce both l_i and l_j below the value for each carrier alone, so that the addition of a further type of excitation does not always increase κ .

Most attention has been paid to the interaction between electrons and lattice vibrations in metals and semiconductors, since this also governs the electrical resistivity and the thermoelectric power. Other carriers of heat which have been considered include spin waves and excitons.

In metals, the ideal electronic thermal conductivity, in conjunction with the ideal electrical resistivity, yields information about the electronic band structure and the electron-phonon interactions. The lattice thermal conductivity, being limited by the same interactions, yields further information about their nature. The lattice thermal conductivity of most pure metals is too small to be measured directly, but can be deduced from measurements on alloys.

In most insulators and many semiconductors, lattice waves are the sole carriers of heat. The intrinsic thermal resistance is due to the mutual interaction of phonons. The theory of this resistance is difficult and its conclusions mainly qualitative. Scattering of lattice waves by crystal imperfections can produce additional thermal resistance, particularly at low temperatures. The temperature variation of this resistance depends upon the nature of the imperfection, so that imperfections can sometimes be identified and studied from the thermal conductivity of insulators and from the lattice thermal conductivity of alloys. The imperfection-induced changes in the electronic thermal conductivity, however, are simply related to the corresponding changes in the electrical conductivity by the Wiedemann-Franz law, and thus do not yield new information.

5.2. Electronic Thermal Conductivity of Metals

The electronic thermal conductivity of metals is closely related to their electrical conductivity. The electron distribution function f ($f = f^0 + g$, wherein f^0 is the equilibrium distribution) is given by a Boltzmann equation of the form

$$\frac{df^0}{dE} v \cdot \left[eF - \frac{d\epsilon}{dT} \nabla T - K \epsilon \nabla T \right] = \frac{df}{dt} \quad (5.2)$$

in the presence of an electric field F and a temperature gradient ∇T . Here v is the electron velocity, E the energy, ζ the Fermi energy, K the Boltzmann constant, $\epsilon = (E - \zeta)/KT$, and e the electronic charge. The right hand side denotes the rate of change of f due to interaction processes. It is always possible to define a relaxation time τ by $df/dt] = -g/\tau$, where g is the deviation from equilibrium; in general τ is a functional of the ratio $F/\nabla T$.

On choosing an electric field F such that the electric current vanishes, the Boltzmann equation (to lowest order in KT/ζ) takes the form

$$\left[\frac{df}{dt} \right] \propto \epsilon \quad (5.3)$$

whereas in the case of electric conduction ($\nabla T = 0$)

$$\left[\frac{df}{dt} \right] = \text{constant}. \quad (5.4)$$

If the solution of Eq. (5.3) leads to the same effective τ as the solution of Eq. (5.4), and in consequence of the high degeneracy of the electron gas, then the resulting thermal conductivity κ and electrical conductivity σ are related by the Wiedemann-Franz law

$$\kappa/(\sigma T) = L = (\pi^2/3) (K/e)^2 \quad (5.5)$$

and L is the Lorenz number. This relation holds for all band structures.

In order that the effective τ be the same for Eq. (5.3) and (5.4), the scattering processes should be elastic or nearly elastic (change in E less than KT). Now

$$\left[\frac{df}{dt} \right] = \left[\frac{df}{dt} \right]_e + \left[\frac{df}{dt} \right]_i \quad (5.6)$$

wherein the first term is due to elastic electron scattering by imperfections and the second is due to inelastic electron-phonon interactions. Correspondingly, $1/\sigma = \rho = \rho_0 + \rho_i$ and $1/\kappa = W = W_0 + W_i$.

For all metals we thus have $W_0 = \rho_0 LT \propto T^{-1}$ and at high temperatures ($T > \theta$, θ being the Debye temperature) $W_i = \rho_i/LT$. Since at high temperatures $\rho_i \propto T$, W_i is independent of T , except for terms of higher order in KT/ζ . At low temperatures $df/dt]_i$ is due to highly inelastic scattering and $\rho_i/W_i T < L$. It can be shown in general that $W_i \propto T^2$ as $T \rightarrow 0$.

These general conclusions are borne out by experiments with few exceptions (Klemens 1956). Thus low temperature thermal conductivities can usually be analyzed (Rosenberg 1955) in the form $W = AT^{-1} + BT^2$, where $A = \rho_0/L$. Rosenberg (1955) has

found several metals where $A < \rho_0/L$ (Ti, Zr, V, Mn, Ce), but it is not certain whether this and similar cases could not be explained in terms of an additional lattice conductivity (Kemp, et al., 1956). There is considerable difficulty in separating the lattice from the electronic component, but by studying different samples and the effect of magnetic field such a separation could perhaps be achieved. There are also anomalies related to anomalies in the low temperature electrical resistivity, for example in magnesium (Rosenberg, 1955; Webber and Spohr, 1957), and this phenomenon is not yet understood. In some cases W_1 does not vary as T^2 at low temperatures (Zn, Cd), but this is related to a strong departure of the lattice wave spectrum at low temperatures from the Debye model, as pointed out by Rosenberg (1957).

There is considerable lack of knowledge about the behavior of $\kappa/\sigma T$ of metals at high temperatures. If $T > \theta$, $\kappa - L\sigma T$ should approach the lattice thermal conductivity κ_g , which should vary as T^{-1} . There has as yet not been any detailed investigation of the high temperature behavior of κ . Experimentally this is a difficult region.

The Bloch equations (5.3) and (5.4) have been set up and solved in the case of complete spherical symmetry only, with spherical energy contours in momentum space and scattering probabilities independent of the initial direction of the electron (the latter restriction excludes Umklapp-processes). The solution of Eq. (5.4) yields $\tau(\epsilon)$ almost independent of ϵ , and a power series in ϵ is a good trial function for a variational treatment, as that of Sondheimer (1950). In the case of Eq. (5.3), however, $\tau(\epsilon)$ is more complicated, unless $df/dt]_0$ predominates, and the Sondheimer method converges slowly. Using a numerical solution (Klemens, 1954) of Eq. (5.3) for the case $T \rightarrow 0$ and $df/dt]_0 = 0$, and interpolating between that and Sondheimer's solution, one can estimate the theoretical thermal conductivity as a function of T and W_0 to within a few per cent. One property of the solution, arising from the ϵ -dependence of τ , is a large deviation from the thermal analog of Matthiessen's rule (W_1 depends on W_0 in the T^2 region and can vary over a range of about 50%); such departures from Matthiessen's rule have indeed been observed (White, 1953; Powell et al., in press).

It is possible to give a qualitative discussion of the Bloch equations. The expression $df/dt]_0$ contains terms due to the change of direction of the electron and terms due to its change of energy. In the case of Eq. (5.4), the latter terms vanish, and the electrical resistance arises from changes in direction of the electron. At

low temperatures this is a many-step diffusion process over the Fermi surface. In the case of Eq. (5.3), both terms contribute, and the terms due to energy changes become dominant at low temperatures. Thus the ratio $\rho_1/W_1 T$ at low temperatures is a measure of the effectiveness of the many-step diffusion process per single step. It varies as T^2 and inversely as the square of the radius of the Fermi sphere.

One would expect $\rho_1/W_1 T$ at low temperatures to depend upon the shape of the Fermi surface. This ratio is substantially in excess of the theoretical value for a spherical Fermi surface in the case of all the monovalent metals, so that Umklapp-processes must be important still at low temperatures, and in all monovalent metals, except possibly sodium, the Fermi surface comes close to or even touches the zone boundary (Klemens, 1954), as is also suggested from other considerations (Cohen and Heine, 1958).

All such conclusions must remain somewhat tentative as long as we have no quantitative solution for Eqs. (5.3) and (5.4), or for the ratio of the τ 's of Eqs. (5.3) and (5.4), except for the case of spherical symmetry. Solutions for other cases would be difficult to obtain, but a few numerical solutions of simple geometry (sphere or simple polygons touching the zone boundary) would be very useful. On the experimental side it would be of interest to study purer specimens, so that ρ_1 and W_1 could be determined at lower temperatures to find evidence for a possible "freezing out" of Umklapp-processes. Substantial improvements in purity would be required.

In the case of polyvalent metals the band structure is more complicated and frequently unknown, so that we lack even a rough understanding of the temperature dependence of ρ_1 . However, W_1 is determined mainly by single step processes at low temperatures, so that it conforms more uniformly to a T^2 dependence. In the case of transition metals, where inter- and transitions should yield components of ρ_1 and W_1 related by the Wiedemann-Franz law, it has been possible to draw limited conclusions from the behavior of $\rho_1/W_1 T$ (Klemens, 1956; White and Woods, 1959).

The temperature dependence of W_1 from low to high temperatures is governed by the ratio of W_1 at low temperatures to ρ_1 at high temperatures. The latter, although including the effects of Umklapp-processes and dispersion of lattice waves, is not sensitive to the finer details of the band structure, nor is W_1 , so that the ratio $W_1(T)\theta^2/T^2W_1(\infty)$, $T \ll \theta$, is fairly constant from metal to metal (Klemens, 1956; Rosenberg, 1955). Here again our lack of knowledge of the band structures and the nature of the electron-

phonon interaction for high-frequency phonons prevents a detailed and quantitative understanding.

The absolute magnitude of all resistivities depends upon the magnitude of the electron-phonon interactions. Since W_1 at low temperatures is less sensitive to details of the band structure and Umklapp-processes than ρ_1 , theoretical values of the interaction constant are best compared with observed values of W_1 at low temperatures, a fact which is as yet not always appreciated. Another direct measure of the electron-phonon interaction is provided by the lattice thermal conductivity.

5.3. Lattice Thermal Conductivity of Dielectric Solids

In a regular crystal with harmonic forces, the lattice vibrations can be resolved into travelling waves

$$u(\mathbf{x}) = \sum \mathbf{b}(\mathbf{q}) \mathbf{p} \exp i(\mathbf{q} \cdot \mathbf{x} + \omega t), \quad (5.7)$$

\mathbf{x} being a lattice point, u the displacement, and the summation is over all wave-vectors \mathbf{q} and polarizations \mathbf{p} ; the direction of \mathbf{p} may be calculated from the interatomic force constants. These lattice waves are quantized, possible values of the energy of each wave increasing in steps of $\hbar\omega$; these quanta of vibrational energy are termed "phonons." The number of phonons in thermal equilibrium, \mathcal{N} , is given by a Planck distribution, and the lattice wave spectrum determines the specific heat (de Launay, 1956).

In a temperature gradient, the phonon distribution \mathcal{N} departs from equilibrium and is determined by the balance between the action of the temperature gradient and the processes tending to restore thermal equilibrium. Thus, the Boltzmann equation for each normal mode (\mathbf{q}, \mathbf{p}) becomes

$$\left[\frac{d\mathcal{N}}{dt} \right] - \mathbf{v} \cdot \nabla T \frac{d\mathcal{N}}{dT} = 0 \quad (5.8)$$

where \mathbf{v} is the group velocity and $d\mathcal{N}/dt$ the rate of change due to interaction processes. The latter may be described formally by a relaxation time τ , i.e.,

$$d\mathcal{N}/dt = (\mathcal{N} - N)/\tau = -n/\tau \quad (5.9)$$

and τ is not only a function of (\mathbf{q}, \mathbf{p}) but in general also a functional of $n(\mathbf{q})$ of all other modes (Klemens 1956, 1958). Formally, $n(\mathbf{q})$ is determined by Eq. (5.8) so that $n \propto \tau$. The thermal conductivity (in an isotropic case) becomes

$$\kappa = (1/3) \int S(\omega) v l(\omega) d\omega \quad (5.10)$$

where $l = v\tau$ (mean free path) and $S(\omega) d\omega$ is the contribution to the specific heat from lattice modes of frequency ω , $d\omega$.

While κ depends upon $S(\omega)$ it also depends on τ , and since the theory of the effective value of τ (or l) is much less certain than the theory of lattice specific heat, it is usually sufficient to base discussions of the lattice thermal conductivity on the Debye model of lattice waves.

Energy is interchanged between the modes (i.e., phonons are created and destroyed) because of static lattice imperfections and because the lattice forces have anharmonic components. Static imperfections give rise to elastic scattering of phonons (from mode q to q' so that $\omega = \omega'$), and the corresponding rate of change of N can be described by a relaxation time $\tau_0(\omega)$.

The interactions due to anharmonicities are processes in which three or more normal modes participate; cubic anharmonicities lead to three-phonon processes (Peierls, 1929), quartic anharmonicities to four-phonon processes. Little attention has been paid to the latter; three-phonon processes are probably more numerous except perhaps at very high temperatures, but there is some experimental evidence (Stuckes 1957) that four-phonon processes are important in indium antimonide. In three-phonon processes a phonon each from two modes (q and q') are created and a phonon from mode q'' is destroyed, or vice versa, such that energy is conserved

$$\omega + \omega' = \omega'' \quad (5.11)$$

and the interference condition

$$q + q' = q'' + 2\pi b \quad (5.12)$$

is satisfied, where b is an inverse lattice vector. (Note that $u(x)$ is invariant under the transformation $q \rightarrow q + 2\pi b$.) The rate of change of $N(q)$ due to such a process contains the factor

$$(N + 1)(N' + 1)N'' - N N' (N'' + 1). \quad (5.13)$$

Processes such that $b = 0$ conserve $\sum q N(q)$ (or quasi-momentum) and are called normal processes (N-processes), processes such that $b \neq 0$ are Umklapp processes (U-processes).

As $N = \mathcal{N} + n$ and as (5.13) vanishes in equilibrium, dN/dt contains terms in n , n' , and n'' , and it is not possible to define a unique relaxation time for three-phonon processes. One can, however, define a relaxation time (τ) for the special case when all n 's vanish except for $n(q)$ of the one mode considered. This relaxation time, for N-processes alone, has been evaluated approximately for

longitudinal (Landau and Rumer, 1937) and for transverse (Pomeranchuk, 1942) phonons.

In general, however, $dN/dt]N) \neq n/\sigma$, and $n(q)$ —or, in view of Eq. (5.9) and Eq. (5.8), $\tau(\omega)$ —is given by the solution of an integral equation. The main difficulty arises from the N-processes, for U-processes play a role similar to the elastic scattering processes and an effective relaxation time τ_U can be ascribed to them. Thus

$$\frac{1}{\tau} = \frac{1}{\tau_0} + \frac{1}{\tau_U} + \frac{1}{\tau_N} \quad (5.14)$$

but τ_N is not known and $\tau_N \neq \sigma$. From (5.13) it follows that $1/\tau_N = 0$ if $\tau(\omega)$ is independent of frequency, so that N-processes tend to make $\tau(\omega)$ independent of ω .

The integral equation (5.14) with the explicit expression for $1/\tau_N$ has not yet been solved, but the following approximations have been proposed. Klemens (1951) has suggested that

$$\begin{aligned} \frac{1}{\tau_N} &= \frac{1}{\sigma} \left(1 - \frac{\tau(\omega_1)}{\tau(\omega)}\right) \text{ if } \omega < \omega_1 = KT/\hbar \\ \frac{1}{\tau_N} &= 0 \quad \text{if } \omega > \omega_1. \end{aligned} \quad (5.15)$$

Thus κ depends on τ_0 and τ_U , low frequency convergence difficulties are avoided, and it is almost independent of σ within rather wide limits. Leibfried and Schloemann (1954) have assumed $\tau(\omega)$ to be independent of ω and have obtained its magnitude by a variational method (in this case equivalent to balancing the gain in quasi-momentum of the phonon gas against the loss due to U-processes). Their choice of trial function implies that $1/\tau_N = 0$, so that N-processes need not be considered explicitly. This is justified only if $1/\sigma \gg 1/\tau_U$, but in their case of combined N- and U-processes their procedure causes little error, because τ_U does not vary very strongly with frequency. However, when the same method is applied to scattering by imperfections, as was done by Ziman (1956), and if $\tau_0(\omega)$ varies strongly with frequency, $\tau(\omega)$ will not be very independent of ω (unless N-processes are infinitely strong), and the method may lead to erroneous results. Thus in the case of scattering by point defects ($\tau_0 \propto \omega^{-4}$), Ziman's method yields a resistivity greater than that given by Eq. (5.15) by a factor ~ 50 .

The best approximation to date is due to Callaway (1959), who assumed that

$$\frac{1}{\tau_N} = \frac{1}{\sigma} \left(1 - \frac{\langle \tau \rangle}{\tau(\omega)}\right); \quad (5.16)$$

this is analogous to Eq. (5.15) for $\omega < \omega_1$ but $\langle \tau \rangle$ is an average value of $\tau(\omega)$ assumed to be the one given by the Leibfried-Schloemann-Ziman momentum balance condition. From Eq. (5.14) and (5.16) and (10), momentum balance requires

$$\int S(\omega) \frac{\sigma(\omega) - \langle \tau \rangle}{\sigma(\omega) + \tau_c(\omega)} d\omega = \int S(\omega) d\omega \quad (5.17)$$

where $1/\tau_c = 1/\tau_o + 1/\tau_U$. Substituting Eq. (5.17) into Eq. (5.16) and (5.14), a value of κ is obtained which is sensitive to the ratio τ_c/σ . While this procedure is only an approximation, and probably a bad one for $\omega < \omega_1$, it is superior to Eq. (5.15) for $\omega > \omega_1$, and since the low frequency modes make a small contribution only to κ , the overall error is probably not serious. However, this is a question which needs further investigation.

All these approximate expressions of κ are fairly cumbersome, especially if more than one resistive process is operative. Most discussions of experimental results are therefore based on the further approximation of assuming the overall thermal resistivity to be additively composed of the resistivity due to individual resistive mechanisms. Provided these components of resistivity have widely different temperature dependence and the measurements are available over a wide temperature range, it is possible to separate the overall resistivity into its components.

The temperature variation of the resistivity due to U-processes has been derived by Peierls (1929) for $T \ll \theta$ and $T > \theta$. At low temperatures ($T \ll \theta$), scattering by imperfections leads to a resistivity $W = 1/\kappa \propto T^{n-3}$ if the scattering cross-section $1/\tau_o \propto \omega^n$, and in this way one can distinguish the effects of different imperfections from the observed temperature variation of the thermal resistivity (Klemens, 1951).

The low temperature thermal conductivities of a variety of dielectric solids have now been measured (Berman, 1953), and the results confirm the qualitative features of the theory to the following extent:

The average mean free path increases with decreasing temperature until at very low temperatures it is limited by the external dimensions of the crystal (Casimir, 1938) and $\kappa \propto T^3$. Berman and coworkers (1953, 1955) have quantitatively studied the boundary resistance on diamond and artificial sapphire. The mean free path can be calculated from the external dimensions, if scattering at the boundaries is completely diffuse. Quantitative discrepancies can be ascribed to partially specular reflection (which may depend on ω , so that κ may deviate from a T^3 dependence), and the magnitude and temperature variation of κ were found to depend on

the state of the surface in a manner consistent with theory. If the material is polycrystalline, κ is greatly reduced, and the mean free path is of the order of the grain size (Berman, 1952).

At higher temperatures κ reaches a maximum and then decreases with increasing T . In the latter region the resistivity is either due to point defects ($W_P \propto T$ if $T \ll \theta$) or due to U-processes ($W_U \propto e^{-\theta/aT}$, $a \sim 1$, if $T < \sim \theta/b$). In a number of materials (e.g., quartz, sapphire, diamond, solid helium, lithium fluoride, etc.) it was found that $W(T)$ varies substantially faster than T . This has been taken to indicate that W_U is important, but as a , depending critically on crystal structure and the dispersion of the lattice waves, is at present unknown, this identification of the thermal resistivity with Umklapp resistance is not always certain.

In a number of substances, $W(T)$ was never found to vary very much faster than linearly with T , indicating the dominance of point defects. Slack (1957) and Berman et al. (1956) drew attention to the importance of the natural isotopic mass variation as a significant source of thermal resistivity. Geballe and Hull (1958) measured a mono-isotopic germanium crystal and found a conductivity substantially larger than that of natural germanium.

The thermal resistivity as a tool in the study of crystal imperfections has yet to be exploited, though its possibility has been demonstrated in a few cases. For example, Slack (1957) studied the effect of a divalent impurity in potassium chloride. Dislocations can be readily detected after deformation (Berman and Tirmizi; Sproull, et al., 1959). The effects of neutron irradiation have been studied (Cohen, 1958; Berman 1951; Berman et al., 1955), and qualitative conclusions could be drawn about isolated displacements and displacement spikes. The nature of the imperfections can be stated in favorable cases with confidence, but their quantitative study presents greater difficulties.

The magnitude of the resistivities depends upon the absolute magnitude of τ_0 , τ_U , and σ . These relaxation times can be calculated from second order perturbation theory; scattering by static imperfections has been discussed by Klemens (1955). It is relatively easy to calculate scattering due to mass difference (isotopes, heavy substitutional impurity) as this perturbs the kinetic energy only, and as the masses are always known. The effects of misfit, distortion, and foreign bonds, affecting the potential energy, are more difficult to treat theoretically, and also the relevant physical parameters are rarely well known. These difficulties are particularly pronounced in the case of scattering by static strain fields. Such processes are analogous to three-phonon processes, the static

strain field taking the place of one phonon (Klemens, 1958). As in the calculations of τ_U and σ , the anharmonicities are estimated from thermal expansion data, using the Grueneisen theory. This determines a Grueneisen parameter γ , which is the average overall for excited modes of an analogous parameter referring to a single mode. These latter parameters form the diagonal elements of a matrix, but the anharmonicity coefficients describing scattering are the nondiagonal elements. Clearly it is not correct to replace the diagonal elements by an average value: it is even less justifiable to use this value for the nondiagonal elements. Yet there seems no way out of this difficulty short of a detailed theory of the lattice forces.

For this reason one cannot rely on the magnitude of the theoretical dislocation scattering (Klemens, 1955, 1958). There is evidence that the theory underestimates dislocation scattering in copper alloys by a factor 3—6 (sec. 5.4). In lithium fluoride, however, where changes in thermal conductivity were compared with etch pit counts (Sproull, et al., 1959), the discrepancy was found to be considerably larger. This could be a special feature of the anharmonicities of this material, but Carruthers (1959) has recently suggested that Klemens' treatment, which relates the perturbation to the strain field, is in error in the case of edge dislocations by a term which depends logarithmically on dislocation density. More theoretical and experimental work is required here.

In the case of mass defect scattering the theoretical expression for $1/\tau_0$ is most probably correct, but comparison between theory and experiments depends on obtaining κ from τ_0 in the presence of N-processes, and in this respect the theory is deficient for point defects, as mentioned above. Many pure crystals have a resistivity which, if ascribed to their isotopes, exceeds the theory based on Eq. (5.15) by factors of order 4, as pointed out by Slack (1957); on the other hand, the difference in resistivity between a natural and a mono-isotopic crystal of germanium (Geballe and Hull, 1958) is roughly in accord with the theory based on Eq. (5.15), as is the resistivity of a germanium-silicon alloy (Toxen, 1958). For stronger mass-defect scattering, i.e., copper-gold alloys (Kemp, et al., 1957), the observed resistivity is somewhat too small, and Toxen (1958) remarked on the apparent non-linear variation of W_P with the mean square mass fluctuation (i.e., with $1/\tau_0$). While this effect could arise from contamination by other point defects, particularly apparent in crystals of low W_P , it could also arise from the effect of N-processes on W_P , which varies with the ratio of σ/τ_0 , and this feature would be an outcome of Callaway's treat-

ment. The dependence of W_p not only on τ_0 but also on σ was recently confirmed by Berman (1958), who measured various samples of isotopically enriched lithium fluoride.

The magnitude of the Umklapp-resistance has been treated approximately by Leibfried and Schloemann (1954): W_U depends on γ , but it also varies rapidly with θ . In the exponential region ($T < \theta/6$) it is not possible to give even a rough estimate of W_U , since it is so critically dependent on the detailed zone structure and law of dispersion. For constant crystal structure, the θ -dependence was verified by measurements of solid helium of various densities (Webb, et al., 1952, 1953). At higher temperatures, where $W_U \propto T$, there is an approximate formula, which has been compared with observed conductivities of various substances (Klemens, 1958). The observed resistivities exceed the calculated ones, and the discrepancy is pronounced whenever γ is particularly low, confirming the suspicion that the inadequate treatment of anharmonicities is at fault. It should also be remembered, however, that the theory assumes as simplification an equivalent monatomic lattice. In the alkali halides the discrepancy is the larger, the more the masses of the two ions differ. This is no doubt due to the extra zone boundaries and extra U-processes resulting from the enlarged unit cell. The intensity of these additional U-processes increases with the relative mass difference (Klemens, 1958), but while this explains the qualitative trend, there is as yet no quantitative treatment of this effect.

Two fundamental points, usually glossed over in theory, have not as yet been satisfactorily explained. The boundary resistance seems to be in general accord with Casimir's theory (1938), which is based on the complete absence of N-processes. These processes would tend to change the phonon flow from a Knudsen to a Poiseuille type and change the temperature variation of the boundary resistance (Klemens, 1958), yet this effect has never been observed. The other difficulty concerns the low frequency longitudinal waves: for these waves, in an isotropic medium, σ should vary as ω^{-4} and κ should diverge except for the effect of dislocations and external boundaries in limiting $l(\omega)$. There is as yet no agreement as to why this divergence is prevented: is there really a weak size dependence at all temperatures in a dislocation-free crystal (Geballe and Hull, 1955); do quartic anharmonicities (Pomeranchuk, 1941) prevent the divergence, or does the variation $\sigma \propto \omega^{-4}$ fail to hold in real crystals, because of their anisotropy and degeneracy of the polarization branches (Herring, 1954)?

So far we have considered nearly perfect crystals. Glasses have a much lower conductivity, due to their structural disorder. From Eq. (5.10) an empirical average mean free path can be derived. Kittel (1949) pointed out that, except at very low temperatures, this mean free path is independent of temperature. However, $l \sim 2\pi/q$, so that the concept of lattice waves is of doubtful validity. At low temperatures, however, $l \gg 2\pi/q$, so that this concept is again justified, and perturbation theory may be used. From $\kappa(T)$ at low T one can deduce $l(\omega)$, which in turn yields the Fourier inversion of a structural correlation function (Klemens, 1958). Thus one can obtain information about the structure of non-crystalline solids from their low-temperature thermal conductivities. There exist few measurements, no systematic work has been done, and the theory is only in a rudimentary form. It is, however, already evident that large and unsuspected differences exist in the structural coherence of vitreous silica.

5.4. Lattice Thermal Conductivity of Metals

In addition to the electronic thermal conductivity (κ_e) of section 5.2, metals and alloys have a lattice thermal conductivity, limited by the same resistive mechanisms as in the case of insulators, with an additional resistivity W_E due to the scattering of phonons by the conduction electrons. For a review of κ_e in metals, see Klemens (1956, 1958). At low temperatures $W_E \propto T^{-2}$ (Makinson, 1938). The same interactions which cause W_E also cause ρ_l and W_l . While Makinson expressed W_E in terms of the high temperature value of ρ_l , it is more advantageous to express W_E in terms of $W_l(T)$, $T \ll \theta$, since both these quantities are relatively insensitive to the details of the zone structure.

The ratio $W_E(T)/W_l(T) \propto (\theta/T)^4$ has been evaluated for a spherical Fermi surface assuming either (a) electrons interacting equally strongly with phonons of all polarizations or (b) electrons interacting with longitudinal phonons only, but τ_E (phonon relaxation time for scattering by electrons) being much greater than σ . In either case W_E is independent of σ ; more generally if τ_E differs for longitudinal and transverse waves, W_E will depend upon σ , and the ratio $W_E T^2$ will be different above and below the temperature where $(\tau_E)_{\text{trans}} \sim \sigma$.

In most metals (exceptions: bismuth and some other semi-metals) $W_E \gg W_l$ so that, if W_0 is small, $\kappa_e \ll \kappa_e$ and cannot be observed. By increasing W_0 so that $\kappa_e \sim \kappa_e$, κ_e can be determined if $\rho(T)$ and $\kappa(T)$ are measured on the same specimen. This is achieved by alloying. Generally one can deduce κ_e up to about 90°K for

alloys of residual resistivity $\rho_0 > \sim 1 \mu\Omega\text{-cm}$. It is found in general that $\kappa_g \propto T^2$ at liquid helium temperatures. By measuring a series of alloys of electron concentration bracketing the parent metal, κ_g for the metal can be deduced by interpolation. In the cases of copper (Kemp, et al., 1957; White and Woods, 1955), silver (Kemp, et al., 1956), and gold (Birch, et al., 1958; in press), the magnitude of W_E is in rough accord with coupling scheme (a), i.e. equal interaction with both polarizations. Unfortunately no such measurements have been made to date on alkali metals because of the difficulty of making suitable alloys.

Chari and de Nobel (1959) have found a change in κ_g/T^2 in various silver alloys around 2°K which they interpret as τ_E not being quite equal for both polarizations and $\tau_E \sim \sigma$ around that temperature, so that the two polarization branches are not coupled by N-processes below that temperature. It would be necessary to make further measurements on silver alloys, and also to look for similar effects in copper and gold alloys, before this interpretation can be accepted with confidence, but on purely theoretical grounds this effect does not seem implausible.

The variation of $W_E T^2$ at helium temperatures as a function of composition should reflect the corresponding variation in the density of electron states and in the electronic specific heat. Very strangely, $W_g T^2$ (where $W_g = 1/\kappa_g$) seems to show a minimum in copper and silver alloys at the electron concentration of the pure metal. There are as yet not sufficient measurements of the variation of the electronic specific heat with alloy composition to test whether a similar effect exists there also: only some copper-zinc alloys have been measured by Rayne (1957, 1958). Again, in copper-gold alloys (Kemp, et al., 1957) $W_g T^2$ was larger than in dilute copper alloys, but pseudo-copper (Cu with equal amounts of Ni and Zn) showed a larger electronic specific heat (Guthrie, 1959) than copper, in contradiction to a simple rigid-band model, which predicts a dependence on electron-concentration only.

There is, however, another possible cause for the variation of $W_g T^2$ with composition. Dislocations cause a resistivity $W_0 \propto T^{-2}$, and dislocation resistance cannot be distinguished easily from resistance due to electrons. Deformed specimens have indeed a much lower value of κ_g/T^2 . The results referred to above were made on annealed specimens, but it is not impossible that even in annealed specimens of alloys there remain a large number of dislocations, due to their trapping and inhibition of their motion by the solute atoms. The observed variation in $W_g T^2$ could be accounted for by residual dislocations whose density is typically of

the order of 10% of the density due to severe plastic deformations. Attempts to eliminate these residual dislocations by annealing at very high temperatures (Birch, et al., 1958; in press, Kemp, et al., 1957) were not successful, but produced some decrease in dislocation density.

It is thus still an open question whether the variation in $W_E T^2$ is due to a change in the density of states or due to residual dislocations or, if both factors contribute, how much of this variation is due to either effect. Measurements of the specific heat, the magnetic susceptibility, and κ_g of the same alloys, varying the annealing treatment, might help to clarify this point, as would measurements of the magnitude of the lattice component of the thermoelectric power (Klemens, 1954).

At higher temperatures κ_g decreases with increasing T . Measurements of κ_g in that range are not accurate, but if $\kappa_g \propto T^{-1}$ then in dilute alloys W_g can be identified with W_U . This can be compared with the prediction for W_U/T of Leibfried and Schloemann (1954). As in the non-metals, the observed values of W_g/T exceed the theoretical values by factors ranging from 3 to 5 (Klemens, 1958).

In less dilute alloys, W_g/T at higher temperatures depends on the solute concentration, and W_P due to the scattering of phonons by these solute atoms can be evaluated. In many cases W_P cannot be accounted for by the mass difference, indicating that the lattice is distorted about the solute atom. Quantitative conclusions are, however, not possible at present. If the solute atom is very heavy, the mass difference is the dominant source of scattering. In copper-gold alloys (Lomer and Rosenberg 1959) there was rough agreement with the theory of Klemens, but the observed resistivity was somewhat too small. This may be due to a failure of perturbation theory for the very strong perturbation in this case. Investigations on less concentrated alloys (using heavy heterovalent impurities) would provide a better test of the theory.

The sensitivity of κ_g due to imperfections and the possibility of identifying the imperfections, as in nonmetals, again make thermal conductivity measurements a tool to study imperfections. Unfortunately, because W_E is already large, the sensitivity is not as good as in non-metals; the other disadvantage is that this method is restricted to alloys. The dislocation resistance due to plastic deformation is readily detected. In some cases, notably Ag-Cd and Cu-Zn, point defects due to deformation have also been detected (Kemp, et al., 1956, 1957). The dislocation resistance gives a measure of dislocation densities, but there is a numerical uncertainty, as discussed in Section 5.3. Lomer and Rosenberg (1959)

studied relative dislocation densities as functions of deformation in copper-zinc and calibrated the dislocation density in two cases against a dislocation count by electron-microscopy. They found that the theory (Klemens, 1958) underestimated W_0 (or overestimated the dislocation density) by a factor ~ 6 . Kemp, et al., (in press) compared changes in κ_g on annealing deformed Cu-Zn and Cu-As alloys with the release of stored energy on annealing and concluded that the discrepancy was somewhere between 3 and 5. It is thus possible to estimate dislocation densities in copper alloys with reasonable certainty from the lattice thermal conductivity. It is not certain whether the same correction factor could be used in other face-centered cubic alloys; it certainly would not apply in other crystal structures. In the case of gold-alloys the discrepancy of W_0 from theory is almost twice that in copper and silver alloys. If this is due to the nature of the anharmonicities in gold, the correction factor for W_0 for gold alloys should also be twice that for copper alloys. It would certainly be interesting to repeat a comparison of dislocation densities from κ_g and from other methods in a gold alloy, to see if this expectation is confirmed.

No work has been done yet to study changes in κ_g of alloys on neutron irradiation, analogous to the changes found in some dielectric solids.

5.5. Thermal Conductivity of Superconductors

Thermal conduction in superconductors can be understood only in terms of a theory of superconductivity. Conversely, such a theory must be able to explain, inter alia, the observed thermal conductivities; their experimental study and phenomenological description may perhaps provide a guide to the formulation of such a theory.

The experimental facts can be summarized—see, for example, reviews (Klemens, 1956; Serin, 1956; Mendelssohn, 1955, 1956)—as follows:

As in normal metals, $\kappa = \kappa_e + \kappa_g$. At the transition temperature, T_C , $\kappa_n(T_C) = \kappa_s(T_C)$ suffices, n and s denoting the normal and superconducting state, and below T_C , $\kappa_{en} < \kappa_{es}$ and $\kappa_{gn} > \kappa_{gs}$. The ratio κ_{en}/κ_{gn} depends upon the relative importance of W_0 and W_1 ; similarly κ_{es}/κ_{gs} depends on the relative importance of W_E and W_I , where W_I is the lattice thermal resistivity due to processes other than electron-phonon interactions. A number of experiments, particularly those of Hulm (1950), indicate that

$$W_{on}/W_{os} = f(T) < 1, f(T_C) = 1 \quad (5.18)$$

$$W_{in}/W_{is} = g(T) < 1, g(T_C) = 1 \quad (5.19)$$

$$W_{En}/W_{Es} = h(T) > 1, h(T_C) = 1 \quad (5.20)$$

$$W_{In}/W_{Is} = 1 \quad (5.21)$$

where $f(T)$, $g(T)$, and $h(T)$ are unique functions of T for any one metal and may even be unique functions of T/T_C for a number of superconductors, or even all superconductors.

The above method of analysis is suggested from the two-fluid model and has not been established yet beyond all doubt; however the uniqueness of $f(T)$ has been tested by measuring samples of different W_o . The ratio $g(T)$ has been established to date only for lead and mercury, and these only for a limited range of T below T_C ; with present-day purities W_o is too large compared with W_i at and below T_C for the other superconductors. Evidence on $h(T)$, obtained mainly from alloys, is inconclusive: various experiments did not agree in obtaining a unique function $h(T)$, but the effects of uncontrolled imperfections (producing W_i) seem difficult to evaluate. More experiments on alloys, paying special attention to physical imperfections, will undoubtedly help to clarify the position regarding $h(T)$, but to extend a study of $g(T)$ to other superconductors would require a spectacular improvement in purity which may not be obtained for some time (Sn and Ta seem a little more promising in this respect).

It is very difficult to interpret changes in κ in the region where $W_{in} \sim W_{on}$, especially as κ_{gs} need not be negligible, even though $\kappa_{en} \ll \kappa_{en}$. Only at very low temperatures, where κ_{en} becomes much less than κ_{gs} , does the interpretation become certain again: in that region W_{Es} has become very small and κ_{gs} is limited by W_i —usually by the external boundaries, by grain boundaries or by dislocations. While dislocations can be studied by means of κ_g in non-superconducting metals only in their alloys, measurements of κ_{gs} at sufficiently low temperatures yield information about dislocations in pure metals (Mendelssohn, 1955, 1956; Montgomery, 1958).

One great difficulty in the study of alloys and "hard" superconductors is the occurrence of the intermediate state. These inhomogeneities in state, due to trapped magnetic flux, can cause an additional thermal resistivity, either by reducing the electron mean free path in the normal regions (Hulm, 1953), or by reducing the phonon mean free path in the superconducting regions (Sladek, 1955; Renton, 1955; Lored and Pippard, 1955). On the other hand, some information about the spatial structure of the intermediate state on a microscopic scale can thus be obtained.

The above phenomenological description is based on a two-fluid model (Heisenberg, 1948) and has been used even though it is generally realized that this model cannot be taken literally. It is assumed that in the superconducting state there are two groups of electrons: normal electrons which carry thermal energy, and superconducting electrons. The concentration of normal electrons decreases rapidly with T , and so does κ_{es}/κ_{en} . Normal electrons are scattered in the same way by imperfections as electrons in the normal state (except that the number of accessible states is decreased), and this assumption accounts quite well for the behavior of $f(T)$. The decrease in density of normal electrons interacting with phonons also accounts roughly for the behavior of $h(T)$. However, it is observed that $g(T)$ decreases more rapidly with T than $f(T)$, at least just below T_c . This seems difficult to explain on any reasonable theory, for it implies that n -electrons in the s -state interact with phonons more strongly than do electrons in the n -state. Yet the behavior of $h(T)$ is not in accord with this possibility, as it seems to reflect the variation in the concentration of n -electrons only.

Bardeen, et al., (1957, 1959) have recently extended a theory of superconductivity to thermal conduction. The role of the normal electrons in the superconducting state is taken by electronic excitations. The results of their treatment are generally equivalent to the results of the two-fluid model, and give it a fundamental justification. However, the same difficulty of explaining the behavior of $g(T)$ is also a problem in their theory and remains the major unsolved problem in the thermal conduction of superconductors.

5.6. Thermal Conductivity of Semiconductors

The thermal conductivity of semiconductors presents a great diversity of behavior. In many cases κ_e is negligible and $\kappa = \kappa_g$ is determined by the processes described in section 5.3. Only in a few cases (see below) is κ_g reduced by phonon-electron scattering. The electronic conductivity can be estimated, at least as regards order of magnitude, from a relation analogous to Eq. (5.15). In the case of a non-degenerate free electron gas this becomes

$$\kappa_e/\sigma T = L' = (2.5 + p) (K/e)^2 \quad (5.22)$$

where $p = d(\log \tau)/d(\log E)$, $\tau(E)$ being the effective electron relaxation time. Thus L' is usually not accurately known, depending on the scattering mechanism and the degree of degeneracy,

but can be used to deduce the order of magnitude of κ_e , which turns out to be very small in most cases.

Following experiments on indium antimonide (Busch and Schneider, 1954) and other semiconductors, which indicated that κ_e was much larger than expected from (1), it was pointed out (Thuillier, 1955; Price, 1955) that if there are two bands (or groups of electrons) of different mobility, the condition that the total electric current vanishes does not imply that the current due to each band vanishes separately, so that there is an additional transport of heat by "ambipolar diffusion"

$$\kappa = L'T\sigma \frac{\mu_1 \mu_2}{(\mu_1 + \mu_2)^2} \left(\frac{\Delta}{2kT} + p + 2.5 \right)^2 \quad (5.23)$$

where Δ is the width of the energy gap between the two bands and μ_1, μ_2 the carrier mobilities in each band. Price (1957) has given a generalized discussion of the Lorenz number and concludes, *inter alia*, that ambipolar diffusion can occur whenever the Fermi energy falls in a forbidden energy range or one of low density of states; in particular it cannot occur in metals, no matter how complicated their band structure.

Recent developments of thermoelectric devices as heat engines have resulted in finding semiconductors with a low value of κ/σ , requiring the reduction of all mechanisms of heat transfer other than electronic conduction, but including ambipolar diffusion. Work at elevated temperatures has often revealed cases when κ exceeds substantially $\kappa_e + \kappa_g$, κ_g being extrapolated from lower temperatures and κ_e calculated from Eqs. (5.22) and (5.23). Joffe and his group (1956, 1959) have attributed the additional conductivity in some cases (e.g., tellurium) to internal electromagnetic radiation and in other cases to heat transported by excitons (e.g., PbTe and PbSe). Heat conduction by internal radiation is important if the attenuation of infrared radiation is low and the index of refraction high, and phenomenologically this process is well understood. The exciton mechanism is more questionable, as there are at present no reliable estimates either of their contributions to the specific heat, of their velocities, or their mean free paths. Furthermore, it should be remembered that measurements of thermal conductivity at higher temperatures are liable to errors due to external radiation, that these errors are difficult to estimate, and that in at least one case (Stuckes, 1957) recent and improved measurements have failed to confirm any anomaly.

The reduction in lattice conductivity can be attained, using the Leibfried-Schloemann (1954) theory as a guide, either in its

original form, or in one of the alternative but equivalent forms proposed by other authors. The most important variable is θ , as $\kappa_g \propto \theta^3$. The decrease of κ with increasing mean atomic mass in similar structures, noted by Joffe (1956, 1959), is a consequence of the variation of θ with atomic mass. Complicated structures tend to have a lower conductivity, even if perfectly regular, because of the additional U-processes, as discussed in section 3, and a further reduction may be achieved by imperfections. Joffe and his group (1956, 1959) have used random solid solutions of ions of different masses but similar chemical properties to reduce κ_g without substantially decreasing the electron mean free path. Another possible approach, which has not yet been investigated, is the use of amorphous semiconductors, making use of the difference in wavelength of the phonons and electrons.

Although in most semiconductors the free electrons do not significantly decrease κ_g , there is evidence that in heavily doped germanium κ_g is spectacularly reduced at low temperatures. This is attributed to scattering of phonons by electrons in impurity bands (Carruthers and Cochran, 1957). Ziman (1956, 1957) has evaluated W_E in the case of a nondegenerate electron gas; the theory differs from the case of metals, because the thermal phonons can interact only with electrons of very much higher energy, whose number decreases exponentially. In the absence of detailed information about the structure of the impurity bands the theory cannot be tested critically.

5.7. Influence of Other Carriers

As pointed out in the Introduction, relatively little attention has been paid to carriers other than free electrons and lattice waves. The contribution of internal electromagnetic radiation and excitons was mentioned in section 5.6. Further work at high temperatures may easily uncover the influence of other types of excitations, in particular those involving atomic rearrangement. From time to time substances have been measured which showed various anomalies in the specific heat; depending on the mobility of the corresponding excitations, these anomalies are reflected in the thermal conductivity, but detailed theories are lacking. Examples are hydrogen bromide (Eucken and Schroeder 1939), methane (Gerritsen and van der Star, 1942), ammonium chloride (Simson, 1951), and rubber (Schallamach, 1941).

The possible importance of spin waves has been long recognized. In paramagnetic salts there could be an additional contribution (Froehlich and Heitler, 1936; Pomeranchuk, 1941; Akhieser and

Pomeranchuk, 1944), but even in the most favorable cases it would be much smaller than the lattice conductivity except below $\sim 10^{-2}$ °K, where in any case it is not easy to attain thermal equilibrium between the lattice and the spins. Sato (1955) suggested that spin wave thermal conductivity could be more important in the (nonmetallic) ferrimagnetic crystals. This conclusion has been criticized (Douthett and Friedberg, 1958).

That spin waves could also reduce the thermal conductivity was suggested by Rezanov and Cherepanov (1953), who considered the reduction of κ_s due to electron-spin interaction in ferromagnetic metals, a reduction which outweighs the contribution of the spin-waves themselves. The same mechanism would, of course, also affect the electrical resistivity.

The same reduction can also occur as a result of spin-lattice interaction. Douthett and Friedberg (1958) measured the thermal conductivity of ferrimagnetic crystals and the influence of a magnetic field (reducing the spin-wave population) on it. They conclude that spin-waves reduce the conductivity by reducing the lattice thermal conductivity. They critically discuss the theory, which is at present only in a semi-quantitative state, and find rough agreement with their observations.

There is no doubt that further research, both theoretical and experimental, is called for on the influence of spin-waves and other excitations on the thermal conductivity, and that such work would prove most fruitful.

5.8. References

- Akhieser, A., Pomeranchuk, I., 1944, *J. Phys., USSR*, **8**, 216.
 Bardeen, J., Cooper, L. N., Schrieffer, J. R., 1957, *Phys. Rev.*, **108**, 1175.
 ———, Rickayzen, G., Tewordt, L., 1959, *Phys. Rev.*, **113**, 982.
 Berman, R., 1951, *Proc. Roy. Soc.*, **A208**, 90.
 ———, Foster, E. L., Rosenberg, H. M., 1955, *Rept. Bristol Conf. Defects in Crystalline Solids* (Phys. Soc., London), p. 321.
 Berman, R., 1952, *Proc. Roy. Soc.*, **A65**, 1029.
 ———, 1953, *Adv. in Phys.*, **2**, 103.
 ———, Simon, F. E., Ziman, J. M., 1953, *Proc. Roy. Soc.*, **A220**, 171.
 ———, Foster, E. L., Ziman, J. M., 1955, *Proc. Roy. Soc.*, **A231**, 130.
 ———, ———, ———, 1956, *Proc. Roy. Soc.*, **A237**, 344.
 ———, 1958, *Zeit. physik. Chemie*, **16**, 146; private communication.
 ———, Tirmizi, S. M. A., private communication.
 Birch, J. A., Kemp, W. R. G., Klemens, P. G., 1958, *Proc. Phys. Soc.*, **71**, 843.
 ———, ———, ———, Tainsh, R. J., *Austral. J. Physics* (in press).
 Busch, G., Schneider, M., 1954, *Helv. Phys. Acta*, **27**, 196.
 Callaway, J., 1959, *Phys. Rev.*, **113**, 1046.
 Carruthers, J. A., Cochran, J. F., 1957, *Madison Conference Proceedings*; private communication.

- Carruthers, P., 1959, *Phys. Rev. Let.*, **2**, 386; *Phys. Rev.* (in press).
- Casimir, H. G. B., 1938, *Physica*, **5**, 495.
- Chari, M. S. R., de Nobel, J., 1959, *Physica*, **25**, 84.
- Cohen, A. F., 1958, *J. Appl. Phys.*, **29**, 591; *Madison Conference Proceedings* (1957); private communication.
- Cohen, M. H., Heine, V., 1958, *Adv. in Physics*, **7**, 395.
- Douthett, D., Friedberg, S. A., 1958, *Carnegie Institute of Technology Report* No. 3.
- Eucken, A., Schroeder, E., 1939, *Ann. Physik*, **36**, 609.
- Froehlich, H., Heitler, W., 1936, *Proc. Roy. Soc.*, **A155**, 640.
- Geballe, T. H., Hull, G. W., 1955, *Conf. Phys. Basse Temperatures*, Paris, p. 460.
- , ———, 1958, *Phys. Rev.*, **110**, 773.
- Gerritsen, A. N., van der Star, P., 1942, *Physica*, **9**, 503.
- Guthrie, G., 1959, *Phys. Rev.*, **113**, 793.
- Herring, C., 1954, *Phys. Rev.*, **95**, 954.
- Heisenberg, W., 1948, *Z. Naturf.*, **3a**, 65.
- Hulm, J. K., 1950, *Proc. Roy. Soc.*, **A204**, 98.
- , 1953, *Phys. Rev.*, **90**, 1116.
- Joffe, A. F., 1956, *Canad. J. Phys.*, **34**, 1342.
- , 1959, *J. Phys. Chem. Solids*, **8**, 6.
- Kemp, W. R. G., Klemens, P. J., White, G. K., 1956, *Austral. J. Phys.*, **9**, 180.
- , ———, Sreedhar, A. K., White, G. K., 1956, *Proc. Roy. Soc.*, **A233**, 480.
- , ———, Tainsh, R. J., 1957, *Austral. J. Phys.*, **10**, 454.
- , ———, ———, White, G. K., 1957, *Acta Metallurgica*, **5**, 303.
- , ———, ———, *Phil. Mag.*, in press.
- Kittel, C., 1949, *Phys. Rev.*, **75**, 972.
- Klemens, P. J., 1951, *Proc. Roy. Soc.*, **A208**, 108.
- , 1954, *Austral. J. Phys.*, **7**, 64.
- , 1954, *Austral. J. Phys.*, **7**, 70.
- , 1954, *Austral. J. Phys.*, **7**, 520.
- , 1955, *Proc. Roy. Soc.*, **A68**, 1113.
- , 1956, *Hdb. d. Phys.*, vol. 14, p. 198.
- , 1958, *Solid State Physics*, **7**, 1.
- , 1958, *Proceedings, Conf. on Noncrystalline Solids*, Alfred, N.Y. (Sept.).
- Landau, L., Rumer, G., 1937, *Phys. Z. Sowjet.*, **11**, 18.
- deLaunay, J., 1956, *Solid State Physics*, **2**, 219.
- Leibfried, G., Schloemann, E., 1954, *Nachr. Akad. Wiss. Gottingen, Math-physik. Kl. IIIa*, No. 4, 71.
- Lomer, J. N., Rosenberg, H. M., 1959, *Phil. Mag.*, **4**, 467.
- Loredo, S. J., Pippard, A. B., 1955, *Proc. Camb. Phil. Soc.*, **51**, 368.
- Makinson, R. E. B., 1938, *Proc. Comb. Phil. Soc.*, **34**, 474.
- Mendelssohn, K., 1955, *Progr. in Low Temp. Physics*, **1**, 164.
- , 1956, *Canad. J. Phys.*, **34**, 1315.
- Montgomery, H., 1958, *Proc. Roy. Soc.*, **A244**, 85.
- Peierls, R., 1929 *Ann. Physik*, **3**, 1055.
- Pomeranchuk, I., 1941, *J. Phys., USSR*, **4**, 259.
- , 1941, *J. Phys., USSR*, **4**, 357.
- , 1942, *J. Phys., USSR*, **6**, 237.
- Powell, R. L., Roder, H. M., Hall, W. J., *Phys. Rev.*, in press.

- Price, P. J., 1955, *Phil. Mag.*, *46*, 1252.
 ———, 1957, *I.B.M. Journ. Res. and Dev.*, *1*, 147.
 Rayne, J. A., 1957, *Phys. Rev.*, *108*, 22.
 ———, 1958, *Phys. Rev.*, *110*, 606.
 Renton, C. A., 1955 *Phil. Mag.*, *46*, 47.
 Rezanov, A. J., Cherepanov, V. J., 1953, *Proc. Acad. Sci., USSR*, *93*, 641.
 Rosenberg, H. M., 1955, *Phil. Trans.*, *A247*, 441.
 ———, 1957, *Phil. Mag.*, *2*, 541.
 Sato, H., 1955, *Progr. Theor. Phys.*, *13*, 119.
 Schallamach, A., 1941, *Proc. Phys. Soc.*, *53*, 214.
 Serin, B., 1956, *Hdb. d. Phys.*, *15*, 219.
 Simson, C. V., 1951, *Naturwiss.*, *38*, 559.
 Slack, G. A., 1957, *Phys. Rev.*, *105*, 829.
 ———, 1957, *Phys. Rev.*, *105*, 832.
 Sladek, W. J., 1955, *Phys. Rev.*, *97*, 902.
 Sondheimer, E. H., 1950, *Proc. Roy. Soc.*, *A203*, 75.
 Sproull, R. L., Moss, M., Weinstock, H., 1959, *J. Appl. Phys.*, *30*, 334.
 Stuckes, A. D., 1957, *Phys. Rev.*, *107*, 427.
 Thuillier, J. M., 1955, *Compt. Rend.*, *241*, 1121.
 Toxen, A. M., 1958, *Phys. Rev.*, *110*, 585.
 Webb, F. J., Wilkinson, K. R., Wilks, J., 1952, *Proc. Roy. Soc.*, *A214*, 546.
 ———, Wilks, J., 1953, *Phil. Mag.*, *44*, 664.
 Webber, R. T., Spohr, D. A., 1957, *Phys. Rev.*, *105*, 1427.
 White, G. K., 1953, *Proc. Phys. Soc.*, *A66*, 559, 844.
 ———, 1953, *Austral. J. Phys.*, *6*, 397.
 ———, Woods, S. B., 1955, *Canad. J. Phys.*, *33*, 58.
 ———, Woods, S. B., 1959, *Phil. Trans.*, *A251*, 273.
 Ziman, J. M., 1956, *Canad. J. Phys.*, *34*, 1256.
 ———, 1956, *Phil. Mag.*, *1*, 191.
 Ziman, J. K. 1957, *Phil. Mag.*, *2*, 292.

6. ELECTRONIC OPTICAL ABSORPTION

D. I. DEXTER

University of Rochester
Rochester, N. Y.

6.1. Introduction

The basic formalism (Condon and Shortley, 1935) for the treatment of optical properties of matter is founded on a combination of the theory of electrodynamics, as expressed in Maxwell's equations, and on quantum mechanics. Most frequently, the electromagnetic radiation is treated in a classical manner, and the material with which it interacts is described quantum mechanically. The interaction itself can be most simply handled by time-dependent perturbation theory, and a fundamental understanding of optical properties was achieved about three decades ago. Since the title and avowed intent of this report is concerned with

"perspectives," an effort will be made to indicate the current status of, and probable future developments in, the areas now of special importance in the field of electronic optical absorption.

6.2. Gases

For dilute monatomic gases there are in principle no remaining problems. It is believed that with accurate knowledge of energy levels and wave functions, all electronic optical absorption lines could be predicted, as to position, strength, and shape. In practice, with the exception of atomic hydrogen, we do not know how to solve the Schrödinger equation with great accuracy, and the problem has been treated in reverse. The observation of absorption and emission lines has permitted the accurate determination of atomic energy levels (Moore) if not the wave functions themselves, for most atoms.

Considerable work is still going on in the field of molecular spectroscopy. The spectra are difficult to interpret in detail in the complex gases, but in principle they are understood in terms of the electronic vibrational and rotational states of the molecules. Various approximate methods have been developed for the quantitative treatment of these states. Considerable work is going on to improve these techniques and introduce better approximations.

The probability, or strength, of a transition gives useful information about the wave functions of the two levels. Such information may be stated in terms of the absorption cross section for electromagnetic radiation, the emission probability, or the "oscillator strength." This knowledge is useful in many branches of science—biology, chemistry, astronomy, physics, etc. Some, but by no means most, of the prominent transitions in atoms have been characterized as to oscillator strength, and work along this line will doubtless continue in coming years. There has recently been increased interest in solving the Hartree-Fock equations for excited states of atoms and ions, thus allowing the computation of transition probabilities and energy level differences for comparison with experiment (Knox, 1957, 1958; Gold and Knox, 1959).

Line shapes are seldom observed to be as simple as the Wigner-Weisskopf theory would predict for isolated atoms (Condon and Shortley, 1935), but discrepancies can be understood in terms of such mechanisms as Doppler and pressure broadening.

6.3. Liquids and Glasses

The situation is totally different in liquids and glasses. Here one has neither the convenience of being able to treat isolated

atoms, nor the advantage of the powerful techniques associated with the periodic potential of a solid. The environment of every atom is different, and a complex statistical theory for optical absorption is clearly required (Dexter, 1958). For the pure material a satisfactory theory has not yet been developed, and this will no doubt be a problem of considerable interest in the near future. The initial step, namely the determination of the density of one-electron energy states, has been attacked for one-dimensional systems by several workers (Dyson, 1953; Schmidt, 1957; Lax and Phillips, 1958; Ford, 1959; Domb, et al., 1959; Doane, 1959), but little progress has been achieved for real materials. Good measurements on the intrinsic electronic absorption of liquids and glasses are badly needed. Very little has been done in this respect, and effort in this direction would be highly rewarding, particularly if the absorption is measured over a wide range of absorption coefficients.

The developments in the theory of absorption in dilute solid solutions have frequently been applicable to dilute solutions in liquids and glasses, and much of what will be said below about the advances and limitations in that context is pertinent here also. Since the issues are simpler and more pointed in the case of solids, discussion of both the successes and failures will be postponed until later.

In summary of the above remarks, two facts stand out: (1) we know how to treat absorption in dilute monatomic gases, and (2) we recognize our inability to do so in the case of glasses and liquids. For solids the situation is intermediate, and most of this section will be devoted to this topic.

6.4. Solids

6.4.1. Centers

First, let us consider the case of isolated substitutional or interstitial impurities in a crystalline solid (Dexter 1958; Seitz 1954; Pekar, 1953; Dexter, 1958b; Markham to be published; McClure, 1952; Gourary, to be published), and examine the present "state of the art" in treating this important aspect of electronic absorption. This will require a discussion of some of the approximations now in use, and we shall predict some of the future activity to be expected. In some important respects, diffusely scattered impurities resemble a dilute gas, in that they do not interact strongly with each other, and problems of "resonance" may thus be avoided. A great many serious approximations are necessarily made, however, in discussing these discrete imperfections. (The generic

terms "center" and "color center" are frequently applied both to isolated impurities and electrons or holes associated with vacancies in ionic crystals.) In dilute monatomic gases the most serious approximation is the "one-electron" approximation, in which correlations among electrons are ignored except for the "accidental" correlations introduced by the Pauli principle. This problem will doubtlessly continue to receive at least a small amount of attention, particularly in systems involving a very small number of electrons. In discussing an isolated center several additional questionable approximations are introduced in order to make the problem tractable. For example, even in the case of a well-localized center, one clearly cannot ignore interactions with, and hence the positions of, the near neighbors. This means that a wave function describing the center is unseparated, involving not only the electronic coordinates of the center, but also the behavior of the neighboring atoms, ions, or molecules. This difficulty is alleviated by the use of the adiabatic (Born-Oppenheimer) approximation, which is applicable when the electronic motion is so rapid as compared with nuclear motion that the nuclei may be considered as being at rest at the instant the electronic motion is computed (Markham 1956). The latter computation should be carried out for all possible sets of coordinates of the atoms with which the center interacts, but in practice it is common to determine the electronic wave functions only for the average nuclear positions and to ignore the effects of motion of the neighboring atoms on the electronic wave function of the center. Considerable attention is being devoted to these approximations (Markham, 1956; Lax, 1952), but more work is needed.

Other related approximations are commonly made, notably the Condon approximation and the use of the Franck-Condon principle (Dexter, 1958b; Lax, 1952; Dexter, 1954). In the former, the dependence of the electronic dipole transition element on the surrounding nuclear coordinates is ignored, and an average is taken symbolically, the average being in practice the value when the neighbors are in their mean position. The Franck-Condon principle is related to the correspondence principle, and is applicable when the final state of a transition includes primarily the highly excited vibrational states of neighboring atomic motion. In many systems, particularly ionic crystals, this is a good approximation (Lax, 1952); in many organic crystals it is poor (Gourary, to be published). Very little is known about how to approach intermediate cases (Klick and Schulman, 1952).

Even with the use of all these approximations the computation

of the parameters of interest in optical absorption of a well-localized center has not been satisfactorily carried out, although from a phenomenological point of view the "configurational coordinate" method of Klick has produced valuable results (Klick and Schulman, 1952; Klick, 1953; Russell and Klick, 1956). Basically the difficulty lies in the circumstance that even if a center is "well-localized," the interactions with neighbors must be considered in detail in computing energy levels and wave functions. (For defect centers, of course, it is the neighbors which *completely* determine the wave functions (Gourary, to be published).) This is a difficult task, although a general prescription for its solution has been sketched (Knox and Dexter, 1956). Much work can be expected on this problem in the near future.

With existing crude methods, it has nevertheless been possible in a few cases to compute with fair accuracy the energy of an absorption band (Williams, 1951) and with rather less accuracy its shape and strength. Further experimental developments can also be expected. Energies of absorption bands are now sometimes measured to 0.1 percent or better, which is adequate for foreseeable purposes, but knowledge of band shapes and strengths of considerably improved precision might well be useful. For example, few oscillator strengths are known to within 10 percent, and factors of 2 or 3 are more common in those few cases for which any information exists. [The difficulty of accurately determining oscillator strengths is related to the circumstance that one would like to keep the center isolated, i.e., the concentration low, to prevent complications associated with the aggregation of the center in question. This tends to keep the "signal" low, in whatever technique one uses. Methods that have been used are the measurement of the integrated absorption coefficient and the determination of the concentration of the center: (1) by chemical methods, (2) by the change in the static magnetic susceptibility of a crystal upon introduction of the centers, (3) by the absolute strength of paramagnetic resonance absorption by the centers, (4) by measurement of radiative transition probability, and (5) by comparison with known oscillator strengths for other centers, assuming a given efficiency for certain photochemical reactions among the different centers.] Band widths at half-maximum are known to 1 percent accuracy in some cases, but the detailed shape over two or three decades of absorption coefficient is not available. This information would be useful in testing some of the theoretical approximations referred to above.

Connected with the absence of sound experimental knowledge

of band shapes and strengths is the difficulty of determining the spatial arrangement of the impurity centers. The theorist is almost forced to presume he is dealing with an isolated center. Experimental evidence is accumulating that the centers are in some cases not randomly dispersed, but clustered. It seems inevitable that this behavior will represent an increasing serious difficulty in the future, and increased attention to this complex phenomenon can be expected.

An outstanding experimental development in the past few years has been the use of techniques involving polarized light in the investigation of optical properties of solids, luminescence and reflection as well as absorption (Feofilov, 1953; Klick and Compton, 1958; Lambe and Compton, 1957; Dutton, 1958; Piper, et al., 1959; Ueta, 1952). With the use of such techniques it has been possible to determine the symmetries of centers involving more than one lattice site, by the preferential destruction or reorientation of the center or by observation of the luminescence from the center. One recent result goes even further. The Tl^+ impurity ion in KCl, with a 1S_0 ground state, is found at 4°K to luminesce as an oriented dipole oscillator when excited by the absorption of polarized light (Klick and Compton, 1958). This center is present in a cubic environment and has been widely discussed in the "tight-binding" approximation. Its apparent ability to "remember" into which of three equivalent P states it was excited, when, according to the general view, it has momentarily a local temperature of many thousands of degrees Kelvin, may require a reconsideration of this prototype of the tightly bound center. It seems certain in any case that experiments with polarized light will continue to produce valuable information.

At the opposite extreme from the well-localized, tightly-bound impurity is the diffuse center encountered in crystals of high dielectric constant and small effective mass, such as group V or III elements in Ge or Si. Such centers are treated in the "effective mass" approximation, in which the wave function of the bound electron (or hole) is taken to be the product of a function like that of an electron (or hole) at the bottom of the conduction band (or top of the valence band), and a hydrogenic envelope function centered at the impurity whose spatial extent is determined by the dielectric constant and the effective mass tensor of the appropriate band (Lax, 1956). In the determination of energy levels, some work has been done to consider in detail the effects of the "core" of the impurity (Kahn, 1957).

A formal theory for absorption by these diffuse centers has been

developed, but very few specific computations have been performed. Absorption band energies, shapes, and strengths should be calculable on the basis of present theories, and theoretical efforts in this direction seem worthwhile.

Experimental results for the diffuse centers require working in the infrared region, but this has not represented great difficulty. The energies of some transitions are known, but band shapes and strengths are largely uncatalogued. Such information may be forthcoming in the next few years. Its usefulness in the field of semiconductors is apparent, e.g., to determine the degree of compensation in cases where the donors and acceptors have distinctly different energy levels.

Inasmuch as band structures are frequently nonspherical in the dependence of energy on crystal momentum, the wave function of the diffuse impurity need not be spherically symmetrical. This could give rise to optical anisotropy, e.g., dichroism, associated with diffuse centers in hexagonal materials like CdS. Polarized emission may likewise be expected, as in the work of Dutton (1958). The increased use of polarized light in the investigation of diffuse centers may produce valuable information about energy band structures in the host crystals.

One of the pressing theoretical problems is the development of techniques for treating cases intermediate between the "tight-binding" and the "effective-mass" extremes (Herring, 1956). There are transitions in centers which seem adequately described by the above extremes, e.g., transitions in unfilled inner shells in the rare earth elements and in group V elements in germanium, respectively, but most centers of interest are intermediate cases for which no quantitative theory has appeared. A theory capable of handling most actual centers can hardly be a simple one, and detailed numerical computations seem inevitable in every case to which it might be applied. Nevertheless, it would be of value to treat a few actual, existing, typical centers to ensure that they can be understood quantitatively.

In principle, by means of the Kramers-Kronig integral relating the real and imaginary parts of the index of refraction, the absorption due to a center allows the determination of other optical properties such as the changes in the real part of the refractive index and in the reflection coefficient. Associated properties like the Faraday and Zeeman effects are partially determined by the absorptive characteristics, and experiments on both absorption and a magnetic field-dependent effect can give information about parameters like the Landé g -factor and effective mass. Such

experiments are only beginning, and it is to be hoped that in the next few years they will give important information. Unfortunately, in some cases the interpretation of the results may require a detailed theory of the center; in the effective mass limit it may not.

Reference has been made above to an important recent development in the study of centers in solids, namely the use of polarized light. Before closing this section on centers, it may be pertinent to mention another recently developed technique of great value in interpreting the results of optical measurements, viz., electron paramagnetic resonance absorption. This technique, which has been successfully applied to defect centers in alkali halides, can "count" electrons in a center and can aid in determining its symmetry and environment. The combination of optical and magnetic resonance techniques has recently been and will continue to be a powerful experimental procedure (Castner, et al., 1958; Delbecq, et al., 1956).

6.4.2. Pure Solids

The remainder of this section will be devoted to nominally pure solids. The qualifying adverb seems likely to be the subject of intensive and extensive work in the near future. Germanium and silicon can be produced in a form effectively pure enough for most purposes, i.e., electrical purposes; however, impurities, vacancies, and dislocations continue strongly to influence optical and electrical results in most systems, frequently in unknown ways and degrees. Much work has gone into methods of preparation and purification of solid materials, and this work must inevitably continue to promote further understanding of perfect single-crystalline solids.

Optical transitions in pure solids may be divided into several classes. Transitions involving energy levels corresponding to the inner shells of electrons are the subject of X-ray and soft X-ray spectroscopy. Much work has been done in this field (Beeman, 1950), and a great deal has been learned about the positions of the lower-lying energy bands. This work is continuing in several laboratories. The general principles for interpreting the results have been understood for many years, but really detailed interpretations capable of giving more than energy levels are few, and advances in this aspect of theory may be hoped for since a considerable amount of experimental data awaits interpretation. This field, though it is still important, no longer occupies the central position it did two decades ago.

Another type of transition expected in solids is a many-particle effect, in which excitations of the entire plasma of electrons are created, i.e., plasmons (Pines, 1955). Theoretical work along this line has been carried out; experimental work is expected to be difficult, in that it will involve the highly inconvenient wave length region of the far ultraviolet or soft X-ray region. Furthermore, the interpretation may be confused by the appearance of the transitions discussed in the preceding paragraph and by the transitions of outer electrons to be discussed below. Nevertheless, this is an area in which important results may be expected.

The remaining transitions are those involving the outer electrons of the constituent atoms. These are usually treated under the headings of "exciton" and "band-to-band" transitions (Dexter, 1958b). The latter concept usually makes use of the very important one-electron approximation, and the former category, while not implying it, frequently employs it. In the last few years increased theoretical attention has been devoted to the fundamental basis for the one-electron approximation in solids, in which the state of a system of many, i.e., 10^{23} , outer electrons is approximated by a wave function consisting of a product of 10^{23} uncorrelated one-electron wave functions. Considerable understanding in principle has been achieved as to why one can neglect the strong interactions among electrons (Bardeen and Pines, 1955), but much more work remains to compute corrections to the simple theory. This work is basic to most properties of solids, not only optical and electrical properties. Pending the expected improvements in the fundamental theory, the following remarks on the low-energy transitions of pure solids will be based on the simple-minded approach.

In this approximation, and neglecting surface effects, the allowed one-electron states of a periodic system are grouped into quasi-continuous energy bands, separated (perhaps) by forbidden bands in which no states occur. If a band of states is partially but not completely filled, the system is a metal; if the valence band is filled, and if the next higher band is separated by a non-zero energy and is empty, the system is an insulator (or semiconductor, depending on the band gap). Let us restrict ourselves to the latter conditions in the following.

In atomic systems optical absorption can result in excited bound states or in ionized states. In extended solids, transitions may excite an electron from the filled (valence) band to a higher (conduction) band, or may remove it from the crystal entirely, ionizing the solid system. The former, a band-to-band transition, is said

to create an "electron-hole" pair, the empty state in the valence band corresponding to a positive hole, and gives rise to photoconductivity; the latter is responsible for the external photoelectric effect.

In insulators still another type of transition can occur, namely an excitation not of one electron between bands, but of a state of the entire system, an "exciton" state, not included within the one-electron approximation. These low-lying exciton states may be thought of in the effective mass approximation as arising from the coulomb interaction between the electron-hole pair, and the exciton is sometimes looked on as a hole and an electron moving through the crystal much like a positronium atom (Wannier, 1937). In the tight binding extreme, an exciton state may be considered as an excited but not ionized atomic state, or better, as a linear combination of such states centered about lattice sites throughout the crystal (Frenkel, 1931; Peierls, 1932).

In either description the exciton's wave function is characterized by a propagation vector k , with momentum $\hbar k$, thus representing a coherent excitation of the entire system. By forming wave packets, with a small range of k , one may define a group velocity and effective mass; since the exciton is a neutral quantum of electronic excitation energy, it can transport energy but not electrical charge.

Many scores of theoretical papers have been written about excitons (Seitz, 1954; Haken, 1958), but for a variety of reasons no completely satisfactory experiment has been reported demonstrating energy transport by this mechanism. This involves a delicate matter, about which there has been considerable controversy (Session by Rochester Semiconductor Conference, 1959). To demonstrate exciton migration conclusively, it is necessary to prove that an observed effect could not be explained by the conversion of the electronic excitation energy to photons which may be subsequently reabsorbed (Broser, et al., 1958). (One of the subtleties is that in some systems the propagation of the exciton is itself determined by the virtual emission and reabsorption of photons; that is, the exciton's properties are determined by the near-zone, dipole-dipole interactions among the atoms of the crystal (Heller and Marcus, 1951). To show exciton motion it is also necessary to eliminate the possibility that the observed effects might be attributable to resonant energy transfer from a stationary exciton or to free electrons and holes resulting from dissociation of the exciton. In the present situation there is no theoretical reason to doubt the existence of excitons, and no com-

elling experimental evidence in favor of important theoretical implications.

This problem might be put in another way. There is no reason to doubt the existence of energy states of lower energy than the lowest conducting state and, indeed, absorption spectra are frequently interpreted in this way; in such diverse materials as alkali halides, cuprous oxide, germanium, crystals of aromatic molecules, etc., various absorption lines and bands are commonly referred to as "exciton" lines or bands, and little objection is raised. The common feeling of uneasiness about the subject of excitons might be phrased as follows: "We have talked about excitons and referred to exciton absorption lines for over a quarter of a century. When is an exciton unmistakably going to manifest itself in energy transfer?" It is certainly to be hoped that the near future will bring experimental clarification of the important topic.

Transitions corresponding to the creation of excitons may occur in the infrared, close to the onset of band-to-band transitions in materials of high dielectric constant, and can be best understood in the effective mass approximation. At the other extreme, in materials such as solid rare gases or alkali halides, they may appear in the far ultraviolet and should be understandable, one would think, on the basis of the tight-binding approximation. Intermediate transition energies probably imply that intermediate descriptions will be required. Quantitative calculations about excited states in the tight-binding approximation are just beginning to appear (Knox, 1959), and satisfactory intermediate methods have not yet been developed. Future progress may be expected in these directions.

Related to the energies of the transitions producing excitons are the strengths of the transitions. The effective mass approximation tends to be good in materials of high dielectric constant for which the forbidden band gap tends to be small and for which the exciton transition energy is also small. But when the dielectric constant is large, the spatial extent of the exciton's wave function is large, giving a small transition element from the ground state. Thus, oscillator strengths tend to be small in these materials (Dresselhaus, 1957; Elliott, 1957). When the tight-binding approximation is valid, oscillator strengths for the production of excitons should be comparable with atomic line strengths, and in the alkali halides they are (Dexter, 1957). For excitons of intermediate transition energies, intermediate line strengths may be expected. It has been rather uncommon in the study of solids

to be concerned with the theoretical or experimental determination of oscillator strengths, but a tendency is growing to try to make use of this information.

The shapes of exciton lines are not well understood. Until recently no good experimental determinations had been made, over a large range of absorption coefficient. Beautiful experimental determinations are now available in a few cases (Dutton, 1958; Hilsch, et al., 1958), and one may expect theoretical activity in an attempt to interpret them. The shapes are clearly related to the temperature, and presumably to the thermally induced aperiodicities of the crystal. The latter will be discussed further in the following.

Related to line shapes and energy transport is the question of the spatial coherence of the exciton. The usual description of the exciton as a momentum state can hardly be satisfactory if the average distance between scattering events is small compared with its wavelength; however, indications are that this is the case, at least in ionic crystals, and very likely in other materials as well. More attention should be devoted to this aspect.

Just as was the case with isolated centers with noncubic symmetries or environments, the use of polarized light may be expected to yield valuable information in optical absorption measurements in noncubic crystals. For the detailed interpretation of these expected results which have begun to appear (Dutton, 1958; Piper, et al., 1959), some generalizations of the current theory must be made, to include the tensor nature of the dielectric constant, for example.

We now turn to a discussion of band-to-band transitions in a periodic potential. Strictly speaking, it is improper to divorce exciton and band-to-band transitions, since the coulomb potential associated with the existence of excitons ensures that the absorption coefficient at the onset of band-to-band transitions must be nonzero. For materials in which the exciton lines have been observed to be pronounced, for example in the alkali halides, the widths of these lines are large, and the band-to-band transitions are not clearly identifiable. At the other extreme, i.e., germanium, the band-to-band transitions are much more noticeable in absorption than are the exciton transitions, and it is operationally significant to discuss band-to-band transitions by themselves.

Solely on the basis of density of states in a static periodic lattice, it is easy to show that the absorption coefficient associated with band-to-band transitions must vary near the onset of absorption as $(E-E_0)^{1/2}$, where E_0 is the minimum vertical energy separation

between the two surfaces of energy versus k , if the symmetries of the valence and conduction band wave functions are such as to make these transitions allowed. It might be mentioned that, depending on the details of the particular band structure, this $(E-E_0)^{1/2}$ dependence may approach arbitrarily near to a step function. The square root dependence, as is the case with the other dependences mentioned below, is only applicable close to the onset of absorption. If the symmetries are such that these transitions are forbidden (in the atomic sense), the absorption coefficient may be expected to vary with energy as $(E-E_0)^{3/2}$.

On the other hand, if we admit the existence of lattice vibration, so that we need not be bound by the conservation-of-momentum selection rule on the electron system alone, we may expect other variations of absorption coefficient with energy in cases where the minimum in the energy of the conduction band surface is not at the same value of k as the maximum of the valence band energy surface. For symmetry-allowed transitions, the absorption coefficient should vary approximately as $(E-E_g)^2$, where E_g is the (lowest, nonvertical) forbidden band gap, and if the transition is not allowed by the selection rules on symmetry, as $(E-E_g)^3$. This sort of behavior is qualitatively observed in some materials, but no quantitative theory of absorption has been developed for any sizable range of absorption coefficient.

The strengths of band-to-band transitions represent a field virtually untouched by the theorists, although experimental results are available in a few materials. It is a difficult problem, but progress may be expected.

The characteristic energies E_0 and E_g are the subject of the band structure calculations recently undertaken and currently attracting considerable attention. It may eventually be possible to predict the energy versus k surfaces, but for the near future the experimental determinations of effective mass tensors and E_g will probably be of more assistance to the theorist than vice versa.

Just as discussed above, in noncubic crystals the use of polarized light in absorption measurements will doubtlessly give valuable experimental information about band structure and symmetry. In both theory and experiment increased emphasis can be expected on this technique.

Valuable information is now being obtained on band structures of semiconducting crystals by observation of optical absorption in the presence of strong magnetic fields. The positions of "Landau levels" are measured in this way, thus allowing the determination of effective masses for both conduction and valence bands. Use of

a magnetic field in this and in other ways is a powerful tool which will doubtlessly find increased use.

As the final topic in the discussion of insulators, we consider the circumstance that crystals are in fact not periodic, but are perturbed by imperfections such as vacancies, impurities, dislocations, and, unavoidably, by lattice vibrations and surfaces. Even in perfect crystals the last-named aperiodicities exist and remain at 0°K if zero-point oscillations are taken into account. Accordingly, it will be necessary to develop theoretical techniques for handling deviations from the periodic potential before full understanding of optical properties can be expected. This is a very interesting and difficult problem indeed, and has attracted considerable theoretical attention with little success. In one dimension the problem is tractable by a trick; the trick does not work in two or three dimensions, and it is dangerous to extend too far an analogy between one and three dimensions. One of the physical problems is the following: If a potential is periodic, there is a definite forbidden energy gap in which no allowed states exist. If the potential is not periodic the density of states will not be zero, but will presumably be very small. Just how small is the question, or one of the questions. There is good reason to believe that the density of states does not go to zero at the "average" band gap, but tails off exponentially, or perhaps as a Gaussian or Poisson distribution. But in a macroscopic crystal one is concerned with absorption coefficients of the order $1\text{--}10\text{ cm}^{-1}$, which is only 10^{-5} or 10^{-6} of the strong characteristic absorption in bulk crystals. Thus, a relatively few states associated with the aperiodicities may strongly influence the optically determined forbidden energy gap. (To a considerably lesser extent the same is true of the combined electrical-thermal determinations.)

These new states may also be of great importance in transport properties. For example, CdS seems to be photoconducting wherever light is absorbed, even though the optical absorption appears at low energies characteristic of exciton lines. Perhaps conduction occurs through dissociation of the exciton into the relatively few one-electron states associated with aperiodicities. Many related questions remain unanswered.

Finally we come to the subject of metals, for which the gross optical properties have been understood for many years on the basic free-electron model (Givens, 1958). Experimentally, there are sizable discrepancies among measured optical constants, discrepancies probably attributable at least in part to surface conditions on the samples, e.g., oxide layers or adsorbed impurities.

Some work is under way to measure optical constants in very high vacuum to avoid these complications. Theoretical and experimental work on band structures of metals will allow the prediction of details of the optical constants, not calculable with the free-electron model (Suffczynski, to be published). Recent theoretical work in this area has to do with nonspherical energy surfaces, Fermi surfaces touching Brillouin zone boundaries, and the anomalous skin effect, in which there may exist a nonlocal relation between electric field and current. Considerable progress may be expected along these lines.

As implied above, the understanding of the optical constants of metals is closely related to many sorts of experiments on transport phenomena, which also depend on the band structure. Powerful techniques, such as cyclotron resonance, are being developed which should greatly increase our knowledge about the energy surfaces in metals. In the meantime, it would be valuable to obtain trustworthy determinations of the real and imaginary refractive indices.

6.5. References

- Bardeen, J., and Pines, D., 1955, *Phys. Rev.*, **99**, 1140.
 Beeman, W. W., 1950, Conference on Applications of X-ray Spectroscopy to Solid State Problems, ONR Report.
 Broser, I., and Broser-Warminsky, R., 1958, *J. Phys. Chem. Solids*, **6**, 386.
 Condon, E. U., and Shortley, G. H., 1935, "The Theory of Atomic Spectra," Cambridge University Press.
 Seitz, F., 1940, "Modern Theory of Solids," McGraw-Hill, New York, 1940.
 Castner, T. G., Kanzig, W., and Woodruff, T. O., 1958, *Nuovo Cimento Supplemento*, **7**, 612.
 Dexter, D. L., 1958a, *Nuovo Cimento Supplemento*, **7**, 245.
 Dyson, F. J., 1953, *Phys. Rev.*, **92**, 1331.
 Domb, C., Maradudin, A. A., Montroll, E. W., and Weiss, G. H., 1959, *J. Phys. Chem. Solids*, **8**, 419; *Phys. Rev.*, **115**, 18 (1959).
 Dean, P., 1959, *Proc. Phys. Soc.*, **73**, 413.
 Dexter, D. L., 1958b, *Solid State Physics*, **6**, 353.
 ———, 1954, *Phys. Rev.*, **96**, 615.
 Dutton, D. B., 1958, *J. Phys. Chem. Solids*, **6**, 101; *Phys. Rev.*, **112**, 785 (1958).
 Delbecq, C. J., Smaller, B., and Yuster, P. H., 1956, *Phys. Rev.*, **104**, 599.
 Dresselhaus, G., *Phys. Rev.*, **106**, 76.
 Dexter, D. L., 1957, *Phys. Rev.*, **108**, 707.
 Elliott, R. J., 1957, *Phys. Rev.*, **108**, 1384.
 Ford, J., 1959, *J. Chem. Phys.*, **30**, 1546.
 Feofilov, P. P., 1953, *Dokl. Akad. Nauk, USSR*, **92**, 743.
 Frenkel, J., 1931, *Phys. Rev.*, **37**, 17, 1276.
 Gold, A., and Knox, R. S., 1959, *Phys. Rev.*, **113**, 834.
 Gourary, B. S., *Solid State Physics*, to be published.
 Givens, M. P., 1957, *Solid State Physics*, **6**, 313.
 Herring, C., 1956, "Photoconductivity Conference," p. 81, Wiley, New York.

- Haken, H., 1958, *Fortschritte der Physik*, Bd VI, 271.
 Heller, W., and Marcus, A., 1951, *Phys. Rev.*, *84*, 809.
 Hilsch, R., and Martiensen, W., 1958, *Nuovo Cimento Supplemento*, *7*, 480.
 Knox, R. S., 1957, *Solid State Physics*, *4*, 413 *Phys. Rev.*, *110*, 375 (1958).
 Kristoffel, N. N., 1958, *Akademii Nauk Estonskoi SSR*, No. 7, 85.
 Klick, C. C., and Schulman, J. H., 1952, *J. Opt. Soc. Am.*, *42*, 910.
 ———, 1952, *Phys. Rev.*, *85*, 154; *J. Phys. Chem.*, *57*, 776 (1953).
 Knox, R. S., and Dexter, D. L., 1956, *Phys. Rev.*, *104*, 1245.
 Klick, C. C., and Compton, W. D., 1958, *J. Phys. Chem. Solids*, *7*, 170.
 Kahn, W., 1957, *Solid State Physics*, *5*, 257.
 Knox, R. S., 1959, *J. Phys. Chem. Solids*, *9*, 238, 265.
 Lax, M., and Phillips, J. C., 1958, *Phys. Rev.*, *110*, 41.
 ———, 1952, *J. Chem. Phys.*, *20*, 1752.
 Lambe, J., and Compton, W. D., 1957, *Phys. Rev.*, *106*, 684.
 Lax, M., 1956, "Photoconductivity Conference," p. 111, Wiley, New York.
 Moore, Charlotte E., "Atomic Energy Levels," Circular 467, National Bureau of Standards.
 Markham, J. J., *Revs. Mod. Phys.*, to be published.
 McClure, D. S., 1959, *Solid State Physics*, *7*, 1.
 Markham, J. J., 1956, *Phys. Rev.*, *103*, 558.
 Pekar, S. I., 1953, *Uspekhi Fiz. Nauk*, *50*, 193.
 Piper, W. W., Johnson, P. D., and Marple, D. T. F., 1959, *J. Phys. Chem. Solids*, *8*, 457.
 Pines, D., 1955, *Solid State Physics*, *1*, 367.
 Peierls, R., 1932, *Ann. Physik*, *13*, 905.
 Russell, G. A., and Klick, C. C., 1956, *Phys. Rev.*, *101*, 1473.
 See Session G., *Rochester Semiconductor Conference*, *J. Phys. Chem. Solids*, *8*, 166-195, 540 (1959).
 Schmidt, H., 1957, *Phys. Rev.*, *105*, 425.
 Seitz, F., 1954, *Revs. Mod. Phys.*, *26*, 7.
 Suffczynski, M., *Phys. Rev.*, to be published.
 Ueta, M., 1952, *J. Phys. Soc., Japan*, *7*, 107.
 Williams, F. E., 1951, *J. Chem. Phys.*, *19*, 457.
 Wannier, G., 1937, *Phys. Rev.*, *52*, 191.

7. LUMINESCENCE

JAMES H. SCHULMAN
U. S. Naval Research Laboratory
Washington, D. C.

7.1. Introduction

Luminescent materials are extremely diversified not only with respect to their chemical composition and physical state, but also with respect to the phenomena they exhibit and the uses to which they are put. Under appropriate conditions luminescence effects occur in the gaseous, liquid, crystalline, and glassy states of the majority of inorganic and organic materials. These effects include the rapid or delayed emission of ultraviolet, visible, or infrared

light on excitation by various kinds of corpuscular or electromagnetic radiations, or by the application of electric fields. Extensive use is made of these effects for illumination, for cathode-ray tube displays in radar and television, for radiography and fluoroscopy, and for the detection and measurement of nuclear and other high-energy radiations. From among these varied aspects of luminescent materials, this summary will attempt to emphasize the most general and significant concepts and problems.

The ability to "tailor make" luminescent materials for specific purposes requires a basic understanding of the mechanisms of energy absorption, transfer, dissipation, and emission. The theoretical understanding of these mechanisms, in a general sense, is not inconsiderable and is increasing. It is as yet, however, neither sufficiently fundamental nor detailed to permit the synthesis of luminescent materials from first principles. Recourse must generally be had to the extensive body of empirical data on luminescent systems as the starting point for the development of new materials, and present-day theory serves primarily as a rough guide to direct further experimental trials.

This evaluation of the state of the art is perhaps best explained in terms of the following classification of luminescent systems according to the nature of the physical processes involved:

1. Systems in which absorption and emission of energy occur in the same group of atoms or "center".
2. Systems in which absorption of energy is accomplished by the host material or by one type of "center" while emission takes place in another center, the energy transfer from absorber to emitter occurring without any accompanying charge transport.
3. Systems in which the absorption, transfer, and emission of energy involves the movement of charge carriers.

7.2. Luminescent Centers

Systems of the first type are exemplified by Tl-activated KCl, in which energy is absorbed by the Tl impurity or "activator" center and is emitted by this same center as luminescence. Such a "simple" system may be said to be understood if one can derive theoretically, from the wave functions of the free activator ion and those of the host lattice ions, a prediction of the energies, widths, oscillator strengths, and decay times of the various absorption and emission bands. Since these quantities are strongly dependent on the details of the coupling between the impurity and its surroundings, it is not surprising that the solution to this problem is not yet in hand.

The general features of this type of system have been qualitatively interpreted for many years in terms of a diagram (Fig. 1) wherein the energies of the ground and electronically excited states of a luminescent center are plotted as functions of a single coordinate which specifies the configuration of the atoms or ions comprising the center. In this way of representing a three-dimensional center, such as a Tl ion surrounded by six Cl ions, it is assumed that there is only one important mode of oscillation; i.e., the breathing mode involving radial displacements of the Cl ions, with different vibrational frequencies characteristic of the center in its ground state and in each of its excited states. The "configuration coordinate" is taken as the Tl-Cl separation. The minimum of each curve corresponds to the equilibrium separation for that excited state, and this equilibrium distance, in general, will be different for the different excited states. Optical transitions between the various states occur vertically on the diagram in accordance with the Franck-Condon principle. Within recent years it has been shown that these diagrams must be modified to treat the center as a quantum mechanical oscillator, and that the existence of zero-point vibrations has important consequences. By the use of such "configuration coordinate" diagrams a number of generally observed properties of simple centers have been successfully correlated. These properties are: the increased wavelength (decreased energy) of the emitted luminescent light compared to the absorbed exciting light (the so-called "Stokes shift"); the temperature-independent luminescence efficiency at low temperatures; the rather sudden decrease in luminescence efficiency at some elevated temperature ("thermal quenching"); the temperature-independent exponential time-decay of the emission following cessation of excitation at temperatures where thermal quenching is negligible; the temperature-dependent increase in the rate of this exponential decay of luminescence at temperatures where thermal quenching is appreciable; the approximate Gaussian shape of many absorption and emission spectra and the skewness of others; the sizable widths of the absorption and emission bands even at temperatures approaching 0°K; and the temperature-dependence of these band widths.

The past decade has seen intensive theoretical and experimental efforts directed at making configuration coordinate curves less schematic and more quantitative. The theoretical effort, which has been concentrated on the KCl:Tl phosphor, assumed that the electronic states involved were states of the Tl ion perturbed by the host lattice (Williams, 1951). In order to simplify the treat-

ment, experimental data concerning the host crystal and the free ions were introduced when possible, an approach, which though valuable, is considerably different in viewpoint from a rigorous theoretical analysis of the problem. A critical examination of this work (Knox and Dexter, 1956) has emphasized the role of empirical data in the theory, has noted certain conceptual difficulties in the treatment, and has pointed out that oscillator strengths of the absorption bands are incorrectly predicted by the theory. It has been concluded that no fundamental attack on the problem along quantum mechanical lines can be made because the presently available wave functions of the Tl ion do not include exchange effects. The most recent theoretical study (Knox, 1959) of systems of the KCl:Tl type shows, moreover, that there must be a considerable admixture of electron transfer states with the excited states of the Tl ion, and that a model constructed solely from free activator ion states cannot provide an adequate description of phosphors of this type. The recent discovery of new absorption and emission bands in KCl:Tl (Williams and Johnson 1959; Patterson, 1958) and studies on Tl as an activator in other systems have further complicated the picture. It is clear that the very difficult, fundamental, and important task of computing quantitative configuration coordinate curves of simple KCl:Tl-type phosphors in a theoretically rigorous manner has not been accomplished. This task remains as one of the major challenges in the field of luminescence theory.

A quite different approach to the configuration coordinate representation of luminescent systems involves the determination of these curves from experimental data. The objectives of this approach are less ambitious than the theoretical derivation of the diagram, but offer a definite means of testing the assumptions and approximations involved in this type of representation, of determining some of the parameters that characterize the luminescent center, and of summarizing the optical properties of the center in a concise form. Assuming, as mentioned before, that the center vibrates in a "breathing" mode, that the slope of the final state curve can be taken as constant over a range of coordinates involved in the transition, and treating the initial state quantum-mechanically and the final state classically, very reasonable but not entirely self-consistent values are obtained for such parameters as the vibrational frequencies of the center in the ground and excited states (Patterson and Klick, 1958). It is not entirely clear whether the inconsistencies observed are due to complications from other centers in the systems studied (principally KCl:Tl

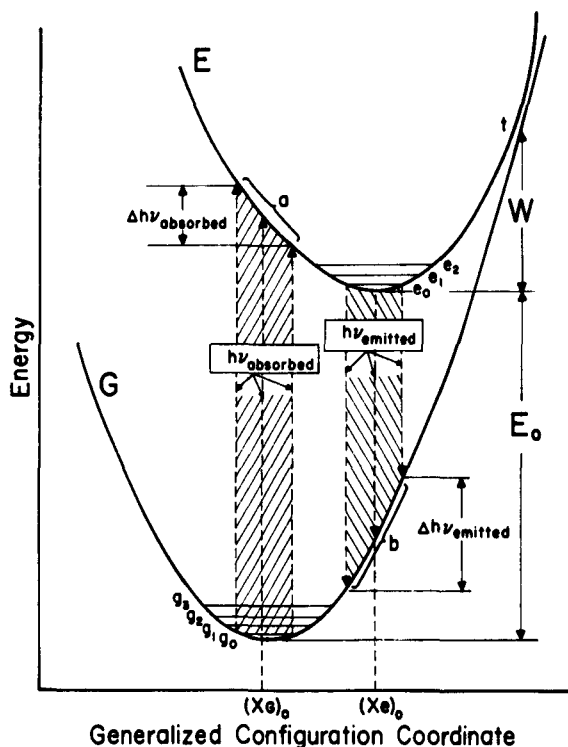


FIGURE 1.—Configuration coordinate representation of a luminescent center.

$g_0, g_1 \dots$ and e_0, e_1 , etc. are the zero-point and higher vibrational states of the ground electronic state G and the excited electronic state E, respectively. Absorptive and emissive transitions occur vertically in the diagram in accordance with the Franck-Condon principle. After absorption of ν light quantum, the center relaxes from the states (a) to the state e_0 , giving up vibrational energy to the lattice. Transitions occur from e_0 to the states (b) with the emission of luminescence, the transition probability being characteristic of the center and independent of temperature. Relaxation of the system to the state g_0 follows, returning the system to its initial state. Stokes Law, $h\nu_{\text{absorbed}} > h\nu_{\text{emitted}}$, is a consequence of these relaxations of the center. The finite width of the absorption and emission bands, even at low temperature, is a consequence of the existence of the zero-point vibrational levels. Radiationless transitions from E to G may occur when the center reaches the point t. The probability of these dissipative transitions is proportional to $\exp(-W/kT)$.

and the F center in various alkali halides), to the assumptions involved in the configuration coordinate representation, or to the approximations made in the treatment. Further work of this type is needed on simple systems that are free of the above-mentioned complications.

The configuration coordinate model treats the interaction of the activator with the lattice as if it were due almost entirely to interactions with the immediate nearest neighbors of the activator. An entirely different treatment of the problem ignores these short-range forces and computes the interaction on the assumption that the lattice is a continuous dielectric in which the activator is embedded (Klick and Schulman, 1957). The results of measurements on F centers in the alkali halides indicate that the important frequency involved is less than that of the long-range longitudinal modes of the lattice which should be the important frequency in this model. On the other hand, notable success has been achieved by means of the continuous dielectric model in predicting the position of the emission peaks of these same centers. Evidence exists to indicate that the longitudinal optical modes become dominant in cases where the center is not strongly localized, such as in materials with high dielectric constants and where the energy levels are close to the band edges. Further investigation of the continuous dielectric model for a variety of systems is desirable.

7.3. Sensitized Luminescence

Luminescent systems of the second general type are exemplified by the calcium halophosphate phosphors which are the predominant luminescent materials used in fluorescent lamps. Phosphors of this type generally contain two impurities that play different roles. The absorption of the exciting light takes place in one of these impurities (the "sensitizer") while the luminescent emission comes principally from the second impurity (the "activator"). In some phosphors the host lattice itself is the absorber or sensitizer, and only one impurity is necessary. When, as in the category of phosphors under discussion, there is no accompanying photoconductivity, the phenomenon is known as "sensitized" luminescence. This effect was first observed many years ago in mixtures of gases. The phenomenon is of importance not only for phosphors involved in lighting, but also for inorganic and organic luminophors used in scintillation counting. The central problem here is to understand the mechanism whereby energy is transferred from the absorber to the emitter, and to relate the transfer efficiency to the properties of the impurity species and of the host material, to

the concentrations of these impurities, and to parameters such as temperature. This problem is considerably less difficult than that of the "simple" centers discussed above because it poses less fundamental questions.

When the sensitizer itself is capable of efficient luminescence emission, and when this emission coincides with a strong absorption band of the activator, the transfer of energy is not necessarily due to sensitized luminescence, but may be due principally to a trivial process of radiative transfer; i.e., emission of photons by the sensitizer and absorption of these photons by the activator. In this case the transfer efficiency is dependent upon the size and shape of the specimen, whereas these factors have no effect on the more interesting process of sensitized luminescence. When both impurities have allowed optical transitions under the foregoing circumstances, it is not always easy to estimate the contributions of the two processes. When the activator, on the other hand, shows weak absorption bands corresponding to forbidden transitions, radiative energy transfer can be ruled out rather easily, and the phenomenon of sensitization can readily be demonstrated.

A fairly adequate general theory of sensitized luminescence has been developed, which involves the transfer of energy by a resonance process due to the overlapping of the electric or magnetic dipole or multipole fields of the impurity atoms or by quantum-mechanical exchange effects (Förster, 1950; Dexter, 1953). Purely classical considerations of coupled oscillators suffice to elucidate the phenomenon, but the problem has also been treated by rigorous quantum-mechanical arguments. Using experimental data concerning the absorption and emission bands and decay times of singly-activated phosphors of the "simple" type previously discussed, the probability per unit time of energy transfer, and hence the transfer efficiency, has been calculated as a function of concentration for sensitizer-activator combinations involving the various types of optically allowed and forbidden transitions mentioned above (Dexter, 1953). These calculations show that the resonance transfer efficiency can be quite high when the sensitizer and activator both have allowed transitions and the overlap between the sensitizer emission and the activator absorption bands is good; transfer occurring with high probability over the order of 10^1 — 10^4 lattice sites surrounding each sensitizer. Less efficient, but nevertheless important, sensitization effects are observable when the activator has forbidden transitions; indeed, the useful and important sensitized inorganic phosphors involve manganese as an activator, which is characterized by a number of spin-flip

and quadrupole transitions. The temperature dependence of sensitization arises from the temperature dependence of the activator absorption and the sensitizer emission band widths, which affect the overlap between these bands. This temperature-dependence is small at low temperatures because the temperature broadening of the bands is negligible compared to their natural widths.

As indicated earlier, the host material may be the sensitizer if it can absorb the exciting light. Such absorption without accompanying photo-conductivity corresponds to exciton production, which may be considered from the point of view of resonant transfer of excitation energy through the crystal from sensitizer to sensitizer. Since the concentration of the sensitizer is very large in the case of host-sensitization, appreciable radiation may be absorbed by the crystal even if the absorption corresponds to a forbidden transition. Sensitizer-sensitizer transfer will then take place by means of exchange rather than by a dipole-dipole process, the speed of exciton propagation in this case will be much lower than that expected when the absorptive transition in the host lattice is an allowed one, and the temperature dependence of the sensitization may be more complicated. Comparatively little work has been done on host-sensitized luminescent systems, and further investigations along this line would be highly important to shed light on the role of excitons in luminescence processes and on the diffusion length of excitons. The usual absence of luminescence from most unactivated "pure" crystals has been interpreted as due to the rapid transfer of excitation energy to imperfections or impurity centers that act as "quenchers" (Dexter and Schulman, 1954). The possible role of resonant transfer processes, long appreciated in the study of quenching effects produced by impurities or by high activator concentrations ("concentration quenching") in luminescent organic solutions, has recently been pointed out for inorganic solid systems (Dexter and Schulman, 1954). Further work is needed on systems of this type with predictable transfer probability to evaluate the importance of this process. Finally, evidence pointing to the contribution of resonance processes to energy transfer in photoconducting phosphors has been put forward (Shinoya, 1956), while other investigations (Melamed, 1958) have shown that this contribution is not likely. Definitive experiments on this point are yet to be performed.

7.4. Luminescence with Charge Transport

Luminescent materials of a third type are exemplified by the highly important class of (Zn,Cd) (S,Se) phosphors. In these

phosphors the luminescence processes appear to be determined more by the host material than by the activator impurities. Energy absorption leads to the movement of charge in the host material. This gives rise to a very important energy transfer mechanism and leads to many phenomena not observable with systems of the two preceding types.

The past decade has seen a remarkable improvement in our understanding of the sulfide-type phosphors based upon a critical examination of the chemistry involved in their synthesis. This has paved the way for a clarification and correlation of their many properties and has led to the prediction of new phosphors. The initial step in this direction came from the questioning of the role of certain previously mysterious addition agents or "fluxes," generally halide salts, used in the synthesis (Kröger and Hellingman, 1949; 1958). It was shown that these are actually incorporated into the phosphor along with the activator, the latter usually being a metallic constituent (generally Cu, Ag, or Au). The role of these fluxes, now more properly called "coactivators," is to compensate for the imbalance of electric charge produced by substitution of a monovalent activator (e.g., Ag) for a divalent constituent of the host material (e.g., Zn). Charge compensation, it was then apparent, could also be accomplished by substitution of trivalent metallic impurities (e.g., Al, In, Ga) along with the monovalent activators (Kröger and Dikhoff, 1950). Not only did this give rise to new methods of phosphor preparation, but it also identified many of the electron trapping states with the electron deficiency introduced by the chemical coactivators. The presence of these traps had previously been deduced from the existence of phosphorescence, thermoluminescence, and infrared-stimulated luminescence. The chemical considerations introduced by the coactivator concept were extended to consider the physical chemistry of phosphor preparation, in general, with particular attention to the gas atmosphere during firing (Kröger, p. 358, Cambridge Symposium, 1955). This has contributed still further to our understanding of the constitution of sulfide type phosphors.

The highly varied luminescence phenomena in these materials have been interpreted rather successfully in terms of the band picture of solids, with the activators and coactivators introducing localized levels in the forbidden region (Fig. 2). However, it must be admitted that much of this success is due to the extreme flexibility of this representation and to the rather free manner in which localized levels with arbitrary properties have been intro-

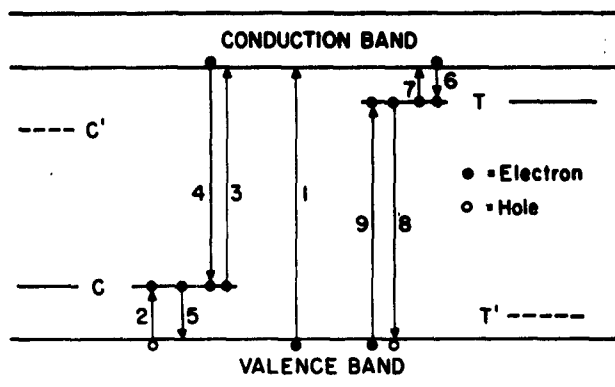


FIGURE 2.—Schematic representation of various possible transitions in a photoconductive phosphor of the ZnS type. (Following H. Klasens, *J. Electrochem. Soc.*, 100, 76 (1953)).

The localized levels, C and T, in the forbidden region originate from chemical impurities (or other defects), and are due either to the impurity itself or to perturbation of the energy bands of the impurity. The C and T levels may have higher excited or "waiting" states, C' and T', respectively. Transitions 3 and 4 below may also involve the C' levels, and transitions 8 and 9 may involve the T' levels. In the unexcited state the C levels are occupied by electrons and the T levels are unoccupied.

- | | |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|
| (1) Fundamental lattice absorption, followed by: | } Excitation processes in Schön-Klasens model. |
| (2) Capture of hole by C level. | |
| Or | |
| (3) Direct absorption in C level. | } Emission process in Schön-Klasens model. |
| (4) Recombination of free electron with hole in C level to give luminescence. | |
| (5) Thermal or optical ionization of hole from C level. | |
| (6) Capture of electron by T level. | } Storage process in Schön-Klasens model. |
| (7) Thermal or optical ejection of electron from T level, followed by transition 4. | |
| (8) Recombination of free hole with electron trapped in T level. | } Phosphorescence, thermoluminescence, or stimulated luminescence in Schön-Klasens model. |
| (9) Absorption filling T level. | |
| | } Direct excitation process in Lambe-Klick model. Indirect excitation in this model is transition 1 followed by transition 6. |

duced as needed to explain the observations. This flexibility is strikingly illustrated by the demonstration that the non-linear dependence of luminescence intensity on excitation intensity can arise from a large family of models characterized by two competing centers in the forbidden region (Duboc, p. S107, Cambridge Symposium, 1955). It is further emphasized by the realization that, in principle, the model generally assumed for the excitation emission cycle is perfectly symmetrical; i.e., it has been proposed (Lambe and Klick, 1955) that the radiative transition in certain sulfide phosphors is the capture of a free hole by a center which has trapped an electron, instead of the capture of a free electron by a trapped hole, which has generally been assumed to be the radiative transition in these phosphors. An important effect of the investigations connected with the above two examples is to remind workers in this field that it is necessary to measure many electrical and optical properties of a system and to study the effect of small alterations in the chemical composition of the material, before concluding that a particular model of the system is unequivocally established. This elementary precaution has frequently been disregarded, particularly in the sulfide phosphors, and the situation has been aggravated by the *ad hoc* postulation of activator, coactivator, and quencher states and distributions of these states. This tendency is rapidly disappearing from the scene, however, and luminescence measurements are being supplemented by the electrical and magnetic measurements commonly involved in the characterization of semiconductors. Progress in this direction is hampered primarily by the present inability to prepare single crystals of sulfide type phosphors which have reproducible properties and contain controlled concentrations of impurities.

Typical of the unresolved problems in this class of phosphors is the current controversy, alluded to above, concerning the nature of the luminescent transition in specific centers in sulfide-type phosphors. It is conceded that the Lambe-Klick model, involving the capture of a free hole by a center which has trapped an electron, is strongly supported by comprehensive, interrelated optical and electrical evidence on certain systems (Lambe and Klick, 1955; Birman, 1959). The applicability of this model to the majority of other centers in sulfide phosphors is considered questionable, however, and for these cases the preference is to invoke the older Schön-Klasens model in which the luminescent transition is the capture of a free electron by a trapped hole. Interpretations of recent measurements on the polarization of the luminescence of CdS and ZnS phosphors may help to decide between these two

views (Birman, 1959). A definitive answer, however, will undoubtedly require the study of each system in question by the combination of electrical, magnetic, and optical techniques mentioned above.

A second typical unresolved problem relates to the models of the activator and trapping centers. Evidence exists that the chemical impurities or the physical imperfections that constitute the activators and the traps are for the most part unassociated. However, since the common activator and trapping impurities are of different charge from the atoms they replace in the crystal and the two types of impurity give rise to centers of different charge in the lattice, a tendency exists for association of the two species. In some cases, where a coactivator-type atom can itself emit a characteristic luminescence (Pr^{+3} in ZnS), this luminescence is different when it performs the function of coactivator in the presence of Ag rather than Cu. An interaction between coactivator and activator type impurities is therefore indicated. With the more usual coactivators, such as the halogens and trivalent ions of the non-rare-earth type, the interaction of the activator and coactivator is not so readily seen; because in contrast to the line emission of rare-earth ions, which leads to easy detection of such interactions, these activators or coactivators yield rather broad emission bands. Nevertheless, small shifts in emission depending on the type of coactivator have been observed in certain cases, and this has been interpreted as evidence for the close spatial association of the two species of defect to form a composite activator center (Prenner and Weil, 1959). "Association" has been proposed both in the above sense, where two differently charged defects agglomerate to form a complex activator center, and in at least two other senses. In one of these concepts of association, the spatial separation of the activator and coactivator species may be quite large; the "association" here involves the assumption that the luminescent transition is from an excited state of the coactivator to the ground state of the activator (Prenner and Williams, 1956; Williams, 1957). In still another context of association involving a spatially closer activator-coactivator pair, the luminescent transition is postulated to be from the ground state of the coactivator to the ground state of the activator (Apple and Williams, 1959). More definitive work on these aspects of "association" is needed, keeping in mind the context in which the term is used.

Even apart from the role of association, entirely satisfactory models for many of the common luminescent centers in sulfide-

type phosphors are still lacking. The so-called "edge emission" in CdS is a case in point. Although all evidence clearly points to this being due to a defect rather than to band-band emission or exciton emission, no good model exists for the center responsible for this emission (Lambe, Klick, and Dexter, 1956). The configurations of other important and experimentally well-known centers, such as the center giving orange emission in un-coactivated high-Cu-concentration ZnS:Cu, have as yet not been identified satisfactorily.

Little work has as yet been done on the band structure of the sulfide-type phosphors, and essentially no quantum mechanical computations have been made concerning the properties of the activator centers or the details of the electronic transitions leading to luminescence in these phosphors. For example, the representation of a center as introducing a definite fixed energy level in the forbidden region may be a considerable oversimplification. It has been pointed out (Schön, 1951) that, due to the rearrangement of the surroundings of the activator upon ionization or excitation, the position of this level will depend on the electronic or thermal process under consideration.

7.5. Electroluminescence

The past several years have seen a very intensive investigation of electroluminescence (Piper and Williams, 1958), the excitation of phosphors by the direct application of electric fields. In the most important case an alternating field is applied to a dispersion of powdered sulfide-type phosphors in a dielectric medium. These studies have been strongly stimulated by the potential commercial importance of the effect as a new type of light source and for the possibility it offers for light amplification and information storage devices. The phenomenon is extremely complicated, and there is disagreement among investigators concerning not only the interpretation of the data but even concerning the data themselves in many cases. In sulfide phosphors the mechanism appears to involve impact ionization of the luminescent centers as the result of acceleration of electrons by the high electric fields established across exhaustion barriers in the phosphor. From the chemistry of the preparation of electroluminescent sulfides, these barriers are believed to be due to the existence of a separate phase of semiconducting Cu_2S in contact with the ZnS. Much remains to be done before the phenomenon achieves its potential practical importance. Investigations based on newly developed techniques for producing thin films of phosphors (Feldman and O'Hare, 1957) may contribute to the realization of this potential. It is probably

a fair statement to say that our fundamental understanding of luminescence processes has been little advanced by the attention given to this complicated phenomenon, and that such advances will come primarily from the research directions mentioned earlier.

7.6. Organic Luminescent Materials

Research on organic luminescent systems has been intensively pursued for many decades (Pringsheim, 1949; Förster, 1950; McClure, 1958; 1959; Wolf, 1959). Since luminescence in these materials is primarily a function of molecular structure, which is the same in the gas, liquid, and solid phases, many experimental approaches have been available in this field that are not available in the study of refractory, insoluble, and non-volatile inorganic systems. Thus, studies on the absorption and luminescence of the gas as a function of pressure; of liquid solutions as functions of the nature of the solvent, the solute concentration, and the viscosity; of crystalline and glassy solid solutions; and of organic molecules adsorbed on substrates, have provided a wealth of data for interpretation. In addition, the special characteristics of organic compounds have made it possible to examine the properties of extended homologous series of compounds, and to prepare derivatives containing a wide selection of substituting functional groups. There has been a parallel growth of the wave-mechanical approach to the theory of organic molecules. These several investigative paths have led to an appreciation of the effects of molecular weight, structure, and bonding on the luminescence, which are exhibited most pronouncedly by cyclic molecules containing conjugated double bonds. Luminescence is promoted by ring closure; by coplanarity of the atoms in the chromophoric group, which eliminates torsional vibrations of the molecule; by the introduction of tightly bound functional groups which cause a shift in absorption to longer wavelengths; by the elimination of loosely bonded substituents with weak bonds to carbon; and by solution of the molecule in a viscous solvent or adsorbing it on a substrate, which inhibits the dissipative processes.

Unfortunately the relationship between work in organic luminescence to that in the inorganic field has not been close, and many concepts and techniques of long standing in the organic investigations have only belatedly been given attention in the inorganic field. One example of this is resonant transfer theory, which was developed in detail for organic systems (Förster, 1950) long before it was applied to inorganic systems (Klick and Schulman, 1957). A second example is the use of measurements on the polarization

of luminescence to gain information concerning the symmetry and orientation of the emitting center, a technique which has been widely used in organic investigations (Förster, 1950) but which has only recently been applied to the inorganic realm (Birman, 1959). Conversely, studies of luminescent organic systems have primarily been concerned with liquid solutions and with pure crystals. The role of crystal imperfections in the luminescence of organic solids has only recently been appreciated and needs much more attention (Wolf, 1959).

As mentioned above, important information concerning organic luminophors has been obtained from studies of these materials in all states of aggregation, not only in the crystalline state. The luminescence of organic liquid solutions, glasses, and plastics is both of fundamental interest and of practical utility. Little attention, by contrast, has been paid by most workers to the luminescence of non-periodic inorganic systems such as glasses. Both direct excitation of an activator and resonant transfer processes can occur equally well in a glassy as in a crystalline host. Furthermore, the environment of an activator in a glass can be varied in an almost continuous fashion by progressive variations in the composition of the glass. Despite a somewhat forbidding chemical complexity, it would appear that research on inorganic non-periodic luminescent systems might yield results of considerable value, both for luminescence and for the elucidation of glass structure.

7.7. References

7.7.1. Reviews and General Articles

- Birks, J. B., 1953, *Scintillation Counters*. McGraw-Hill. New York.
- Fonda, G. R., and Seitz, F. (Editors), 1948, *Solid Luminescent Materials*. John Wiley and Sons, Inc. New York.
- Förster, Th., 1950, *Fluoreszenz organischer Verbindungen*. Vandenhoeck and Ruprecht. Göttingen.
- Forsythe, W. E., and Adams, E. Q., 1948, *Fluorescent and Other Gas Discharge Lamps*. Murray Hill Books, Inc. New York.
- Garlick, G. F. J., 1949, *Luminescent Materials*. Clarendon Press. Oxford.
- , 1950, Cathodoluminescence. *Advances in Electronics*, 2, 151.
- , 1958, Luminescence. *Handbuch der Physik*. XXVI. Light and Matter II. Springer Verlag. Berlin.
- Klick, C. C., and Schulman, J. H., 1957, *Luminescence of Solids*. *Solid State Physics V*. (Seitz, F., and Turnbull, D., Editors). Academic Press, Inc. New York.
- Kröger, F. A., 1947, *Some Aspects of the Luminescence of Solids*. Elsevier Press. Amsterdam.
- , 1956, *Inorganic Crystal Phosphors*. *Ergebn. exakt. Naturw.* XXIX. Springer Verlag. Berlin.

- Leverenz, H. W., 1950, *An Introduction to the Luminescence of Solids*. John Wiley and Sons, Inc., New York.
- McClure, D. S., 1958, *Electronic Spectra of Molecules and Ions in Crystals*. Part I. Molecular Crystals. *Solid State Physics VIII*; 1959, Part II. *Spectra of Ions in Crystals*. *Solid State Physics IX*. (Seitz, F., and Turnbull, D., Editors.) Academic Press, New York.
- Piper, W. W., and Williams, F. E., 1958, *Electroluminescence*. *Solid State Physics VI*. (Seitz, F., and Turnbull, D., Editors.) Academic Press, Inc., New York.
- Pringsheim, P., 1949, *Fluorescence and Phosphorescence*. Interscience Publishers, Inc., New York.
- Proceedings of Cambridge (U.K.) Symposium on Luminescence. Supplement No. 4, *British J. Applied Physics*, 1955.
- Schön, M., and Welker, H. (Editors), 1958, *Semiconductors and Phosphors*. Interscience Publishers, Inc., New York.
- Wolf, H. C., 1959, *Electronic Spectra of Aromatic Molecular Crystals*. *Solid State Physics IX*. (Seitz, F., and Turnbull, D., Editors.) Academic Press, New York.

7.7.2. Original Articles

- Apple, E. F., and Williams, F. E., 1959, *J. Electrochem. Soc.*, **106**, 225.
- Birman, J. L., 1959, *Phys. Rev. Letters*, **2**, 157.
- Dexter, D. L., 1953, *J. Chem. Phys.*, **21**, 836.
- , and Schulman, J. H., 1954, *J. Chem. Phys.*, **22**, 1063.
- Duboc, C. A., 1955, *Proceedings of the Cambridge Symposium on Luminescence*. Supplement No. 4. *British J. Applied Physics*, Paper No. 21, p. S107.
- Feldman, C., and O'Hare, M., 1957, *J. Opt. Soc. Amer.*, **47**, 300, 790.
- Knox, R. S., and Dexter, D. L., 1956, *Phys. Rev.*, **104**, 1245.
- , 1959, *Phys. Rev.*, **115**, 1095.
- Kröger, F. A., and Hellingman, J. E., 1949, *J. Electrochem. Soc.*, **93**, 156.
- , and ———, 1958, *J. Electrochem. Soc.*, **95**, 68.
- , and Dikhoﬀ, J., 1950, *Physica*, **16**, 297.
- , 1955, *Proceedings of the Cambridge Symposium on Luminescence*. Supplement No. 4, *British J. Applied Physics*. Paper No. 11, p. S58.
- Lambe, J., and Klick, C. C., 1955, *Phys. Rev.*, **98**, 909.
- , ———, and Dexter, D. L., 1956, *Phys. Rev.*, **103**, 1715.
- Melamed, N. T., 1958, *J. Phys. Chem. Solids*, **7**, 146.
- Patterson, D., and Klick, C. C., 1957, *Phys. Rev.*, **105**, 401.
- , 1958, *Phys. Rev.*, **112**, 296.
- Prener, J., and Williams, F. E., 1956, *J. Electrochem. Soc.*, **103**, 342.
- , and Weil, D. J., 1959, *J. Electrochem. Soc.*, **106**, 409.
- Schön, M., 1951, *Z. Naturforsch.*, **6a**, 287.
- Shinoya, S., 1956, *Bull. Chem. Soc., Japan*, **29**, 935.
- Williams, F. E., 1951, *J. Chem. Phys.*, **19**, 457.
- , 1957, *J. Opt. Soc. Amer.*, **47**, 869.
- , and Johnson, P. D., 1959, *Phys. Rev.*, **113**, 97.

Part V

DIFFUSION AND MASS TRANSPORT IN SOLIDS

Chairman

Prof. DAVID LAZARUS
University of Illinois
Urbana, Ill.

Members

Dr. R. E. HOFFMAN
General Electric Co.
Schenectady, N. Y.

Prof. H. B. HUNTINGTON
Rensselaer Polytechnic
Institute
Troy, N. Y.

Mr. A. D. LECLAIRE
Atomic Energy Research
Establishment
Harwell
Didcot, Berks., England

Prof. R. A. SWALIN
University of Minnesota
Minneapolis 14, Minn.

CONTENTS

	Page
1. INTRODUCTION	254
1.1. Acknowledgment	257
1.2. Defining Equations	257
2. EXPERIMENTAL TECHNIQUES	259
2.1. Isothermal Diffusion	259
2.2. Electrolytic Conductivity	262
2.3. Concentration and Mobility of Defects	263
2.4. Internal Friction	265
2.5. Special Techniques	266
3. ATOMIC MODELS	267
3.1. Reaction Rate Problem for Diffusion in Solids	267
3.2. Diffusion Mechanisms	271
3.3. Temperature and Pressure Dependence of the Diffusivity	272
3.4. Theoretical Approaches to Atomistic Mechanisms	273
3.5. Influence of Chemical, Electrical, and Thermal Gradients	274
3.6. Diffusion of Impurities	277
3.7. Diffusion in Alloys	279
3.8. Anisotropic Diffusion	280

	Page
4. PHENOMENOLOGICAL AND SEMI-EMPIRICAL MODELS.....	281
4.1. Correlation Between the Diffusion Coefficient and the Melting Point	281
4.2. Correlation Between the Activation Energy and Various Physical Properties of the Crystal	282
4.3. Correlation Between the Diffusion Coefficient and the Activation Energy	283
4.4. Correlations Based on the Elastic Modulus	283
4.4.1. Relation Between Q and Modulus	283
4.4.2. Correlation Between D ₀ and Q	284
4.5. Correlations for Solute Diffusion	285
4.6. Correlations with Concentration and the Effect of Concentration Gradients	286
4.6.1. Phenomenological Treatment Involving Concentration Gradients	286
4.6.2. Correlations with Solute Concentration	286
4.7. Volumes of Activation	287
5. PROBLEM AREAS	287
5.1. Kirkendall Effect and Chemical Diffusion	287
5.2. Relation Between Chemical and Self-Diffusion	290
5.3. Diffusion in Grain Boundaries and Dislocations	291
5.4. The Importance of Self-Diffusion Measurements	294
5.5. Vacancy Configurations: Experiments Under Hydrostatic Pressure	294
5.6. Some Properties of Defects Measured in Equilibrium: Lattice Expansion	297
5.7. Properties of Defects Measured Under Non-Equilibrium Conditions	297
5.8. Interactions of Vacancies With One Another: Vacancy Pairs ..	301
5.9. Interactions of Vacancies With Impurities: Impurity Diffusion	302
5.10. Correlation Effects: Mass Dependence of Jump Rates	304
5.11. Conclusions	306
6. REFERENCES	308

DIFFUSION AND MASS TRANSPORT IN SOLIDS

1. INTRODUCTION

There are few phenomena related to solid materials as all-embracing in their scope as diffusion. So many of the familiar processes characteristic of solid-state reactions depend upon the rate of mass flow of atoms within solids that even to list them all would be a considerable task. Many phenomena connected with the corrosion and oxidation of materials are ultimately related to diffusion, as are those associated with precipitation hardening, grain growth, sintering and various other solid-state transformations. To this list we may also add such items as the photographic

process, electrical conduction in ionic solids, high-temperature creep, and annealing of radiation damage, without by any means exhausting the number of well-known, diffusion-limited phenomena.

In the present report we can hardly hope to recount the extensive history of a subject which has been vigorously investigated for the better part of a century. Rather, we shall direct our attention mainly to a consideration of what diffusion studies have told us, and may yet tell us, about the basic properties of solid materials. We shall be concerned chiefly with diffusion in relatively simple solid systems, and shall generally not consider such problems as the diffusion or permeation of gases through solids, surface diffusion, diffusion in liquids, or diffusion in multicomponent and multiphase systems.

The magnitude of diffusional flow in seemingly "impermeable" solids frequently comes as something of a surprise. At first glance, it seems difficult to reconcile geological evidence of the stability of crystalline solids and microscopic observations, which indicate that solids are essentially rigid and perfect lattices of atoms executing small thermal vibrations about their equilibrium positions, with any appreciable amount of mass flow. Yet, indisputable and quantitative measurements show that in most solids, and at temperatures well below the melting point, the mean time of stay of any single atom in its "equilibrium position" may well be less than a microsecond!

It is precisely this dilemma which has motivated the bulk of recent scientific interest in diffusion. What mechanisms can we propose which will allow such enormous mass flow, while simultaneously permitting a crystalline solid to preserve its overt aspect of perfect order and rigidity? Even the most "perfect" lattices must contain some imperfections, resulting either from accidents of growth or other causes. How can the presence of these imperfections be used to explain the observed diffusion rates, or, taking the opposite approach, how can we employ studies of diffusion to elucidate the nature of the imperfections? It is this dual question which forms the basis of the present discussion.

As we shall see, our detailed knowledge of the relationship between imperfections and diffusion is still in a highly rudimentary state. We are quite convinced that diffusional flow is intimately associated with the existence of mobile lattice defects, but are far from being able to provide a precise description either of the fundamental atomic motion in the elementary diffusional jump or of the nature of the defects responsible for mass flow.

In the present report, we shall, after defining a few necessary terms, first examine the current experimental techniques which have proved useful for diffusion studies. We shall consider, among the multitude of known or potential techniques, those which are best suited to extraction of unambiguous data, considering especially the limits to which such results are indeed "unambiguous."

Next, we shall review the theories and atomic models which have been invoked to explain observed diffusion phenomena. We shall attempt to assess our present state of knowledge of this subject, considering the extent to which the theoretical treatments are self-consistent and in agreement with experiment.

In surveying the large body of experimental diffusion data which has been accumulated over the past several decades, many workers have become aware of several distinct correlations between diffusion and other macroscopic properties of solids, most notably elastic behavior and melting. These observations have given rise to a number of extremely useful semi-empirical rules which form an important tool for the experimenter. Some of these rules are intuitively appealing by virtue of their inherent simplicity, and they have frequently been used to support detailed theoretical models with somewhat more vehemence than justified by the basic assumptions or the accuracy of the relevant data. We attempt, in this report, to survey some of the more commonly accepted semi-empirical correlations, and to assess objectively their limits of applicability.

In diffusion studies, as in other areas of materials research, many of the simpler problems have already been analyzed or treated in a satisfactory manner. Characteristically, the results have frequently posed more questions than they have answered. While there is a growing need for accurate diffusion data for all classes of materials, work in a few specific areas may reasonably be expected to make major contributions to our understanding. In the final section of this report, we shall attempt to delineate those areas which, in the opinion of the authors, are most urgently in need of serious theoretical and experimental investigation.

For the purposes of this discussion, no attempt has been made to compile a complete bibliography relevant to the field of diffusion. Wherever possible, the reader is referred to literature cited in one of several recent review articles listed at the end of this report¹, in which extensive bibliographies may be found.

¹ References cited in review articles are designated parenthetically by date and the listed number of the review article, e.g., (1957,R4). Very recent papers not covered in the review articles are listed separately.

1.1. Acknowledgment

In preparing this report, the authors have received valuable advice and criticism from many of their colleagues. We wish to express our appreciation to those who were so helpful, and particularly, for their written comments, to the following: C. E. Birchenall, J. A. Brinkman, G. J. Dienes, J. Friedel, F. G. Fumi, L. A. Girifalco, Y. Haven, C. Herring, R. W. Keyes, L. Himmel, A. B. Lidiard, J. R. Manning, N. H. March, C. W. McCombie, G. W. Rathenau, A. H. Schoen, P. G. Shewmon, C. P. Slichter, L. M. Slifkin, and R. Smoluchowski.

1.2. Defining Equations

Mass transport in solids results from the thermal migration of atoms through the crystalline lattice, either entirely at random, or influenced by the presence of chemical, thermal and electrical gradients. Diffusion is generally measured in terms of a diffusion coefficient, usually designated by the symbol D , which is defined in terms of macroscopic variables by equations analogous to those used to describe heat flow. If we let J_i represent the instantaneous current of atoms of type i induced by a concentration gradient, $\text{grad } c_i$, the diffusion coefficient of the species, D_i , is defined in terms of Ficks' first law as

$$J_i = - D_i \text{grad } c_i. \quad (1.1)$$

Usually the current and concentration gradient are not time-independent, but vary because of the compositional changes accompanying the diffusional flow. In such situations, if the change in concentration gradient with time is determined solely by the current J_i , the continuity condition requires that

$$\text{div } J_i + \partial c_i / \partial t = 0, \quad (1.2)$$

and we may define the diffusion coefficient in terms of the familiar second-order diffusion equation, usually referred to as Fick's second law,

$$\partial c_i / \partial t = \text{div } (D_i \text{grad } c_i). \quad (1.3)$$

This equation is not directly integrable if D_i is a function of position and time, but may be used for the numerical evaluation of a "chemical diffusion coefficient" using the methods introduced by Matano. In special cases, for example where $\text{grad } c_i$ represents only an isotopic gradient, D_i is constant, and the equation is directly integrable, the concentration c_i being given by the well-known error-function solution. The diffusion coefficient then corresponds to the true "self-diffusion" coefficient, and describes the purely random motion of atoms of species i within the lattice.

The self-diffusion coefficient, D_{si} , may be defined exactly from the theory of random flights, in terms of the mean frequency of individual uncorrelated atomic jumps, Γ_i , and the mean jump distance, r , as

$$D_{si} = \frac{1}{6} n r^2 \Gamma_i, \quad (1.4)$$

where n is the number of independent paths which the diffusing atom may follow.

Equation (1.4) has an exact atomistic interpretation, but does not, unfortunately, lend itself easily to direct experimental verification, because of difficulties in direct measurement of the random jump frequency Γ_i . Most accurate experimental methods evolve from situations best described by Fick's Laws. However, Fick's Laws, at least in the form of Eq. (1.1), ignore the fact that the driving force for mass flow is in general a gradient in free energy rather than in concentration, and further, that in a many-component system, the rate of flow of any one constituent may be related to the gradients in chemical potential of the other constituents, as well as to the presence of thermal and electrical gradients. If the currents of mass, heat and electricity are linear functions of the driving forces, we may write generally for a system of m constituents,

$$\begin{aligned} J_i &= - \sum_{j=1}^m L_{ij} T^{-1} \text{grad } \mu_j - L_{ih} \text{grad } 1/T - L_{ie} T^{-1} \text{grad } \phi, \\ J_h &= - \sum_{j=1}^m L_{hj} T^{-1} \text{grad } \mu_j - K \text{grad } T - L_{he} T^{-1} \text{grad } \phi, \\ J_e &= - \sum_{j=1}^m L_{ej} T^{-1} \text{grad } \mu_j - L_{eh} T^{-1} \text{grad } 1/T - \sigma \text{grad } \phi. \end{aligned} \quad (1.5)$$

Here J_i , J_h and J_e are, respectively, the mass flux of the i th constituent, the flux of heat, and of electricity. The μ_j are the chemical potentials, T is the absolute temperature, and ϕ is the electric potential. The L matrix is composed of constants of proportionality. For $T^{-2}L_{hh}$ and $T^{-1}L_{ee}$ we have used the familiar symbols K and σ , the thermal and electrical conductivities. By virtue of the principle of microscopic reversibility, the L matrix is symmetrical. The specification of these coefficients for any material is sufficient to describe its transport properties.

The diffusion coefficient, D , is almost invariably found experimentally to vary with temperature in a simple exponential manner described by the familiar Arrhenius equation,

$$D = D_0 \exp (-Q/RT), \quad (1.6)$$

where the "frequency factor," D_0 , and the "activation energy," Q , are independent of temperature over the range investigated; R is the gas constant, and T is the absolute temperature. When we attempt to describe diffusional flow in terms of atomistic models involving lattice defects, we must consider separately the probability that a defect is present at an appropriate lattice site adjacent to a diffusing atom and that, in this configuration, the atom also has sufficient energy to execute the elementary diffusional jump. If both of these probabilities are governed by typical Boltzmann expressions, we may evidently associate

$$Q = \Delta H_f + \Delta H_m, \quad (1.7)$$

where ΔH_f is the energy required to form the appropriate defect, and ΔH_m is the energy required for the elementary diffusional motion.

The frequency factor, D_0 , involves the product of several terms: the square of the jump distance; a frequency term characteristic of the vibrational frequency of the diffusing atom; a term taking account of the possible correlation between successive atom jumps; and an entropy term which may be temperature dependent, arising from the fact that the lattice vibrational frequencies are changed by formation and excitation of a defect. The exact form of the expression used for D_0 depends on a number of assumptions made regarding the detailed atomic mechanism.

2. EXPERIMENTAL TECHNIQUES

In order to provide the framework for a critical analysis of the state of knowledge in the field of diffusion, it is first necessary to examine the experimental techniques from which this knowledge has been obtained. From among the large variety of techniques available, we have selected for discussion here only those which have received rather wide usage or which possess some desirable element of uniqueness. We shall pay particular attention to the kinds of information obtainable with these techniques and to the accuracy and reliability of the results. No attempt will be made to describe the experimental procedures in detail; for the most part, these have been adequately presented in the literature (Tomizuka, 1959, R6).

2.1. Isothermal Diffusion

Prior to the widespread availability of artificial radioisotopes, nearly all diffusion measurements were limited to the study of mass flow in the presence of appreciable chemical concentration

gradients, and in metal systems this was essentially the only method available for studying mass flow. The discovery and confirmation of the Kirkendall effect, approximately a decade ago, provided clear evidence that the two components of a binary alloy, like those in ionic solids, must be expected in general to diffuse at different rates, and therefore that diffusion in a chemical gradient must be accompanied by some sort of plastic deformation (except, of course, in interstitial alloys). Detailed examinations of the microstructure of the diffusion zone have shown that transverse dimensional changes and porosity are also frequently encountered in chemical interdiffusion experiments.

These developments emphasized the fact that diffusion in a chemical gradient was definitely not the simple, homogeneous, unidirectional flow process that had been assumed in the standard methods for evaluating diffusion coefficients from concentration-distance data. As a result, much of the older data obtained from measurements in rather large gradients are now considered unreliable. Chemical diffusion measurements have today generally been discarded as a direct means of studying elementary atomic jump processes, except under very small concentration gradients.

Fortunately, at about the time the Kirkendall experiments began to cast doubt on the reliability of chemical diffusion coefficients, a supply of a wide variety of artificial isotopes became available; this provided a unique opportunity for studying the mobility of individual atomic species in essentially homogeneous media. Since the elimination of the complicating side effects that accompany diffusion in a chemical gradient greatly facilitates the comparison between theory and experiment, it is not surprising to find that, where feasible, tracer methods are now used almost exclusively in diffusion investigations.

In addition to the inherently high sensitivity and simplicity of tracer techniques, a greater and more widespread recognition of some sources of systematic error has now been achieved, as well as a general improvement in techniques, so that tracer diffusion coefficients can today be measured with a high degree of accuracy. Tomizuka (1959, R6) has recently given a comprehensive and critical discussion of the experimental details and the precautions that are necessary to achieve reliable results. However, it seems worthwhile here to re-emphasize some of the more important factors that frequently lead to a lack of reproducibility in the measurements.

To obtain diffusion coefficients which are characteristic of the bulk material, it is important that special attention be paid to

possible contributions resulting from diffusion along internal surfaces such as grain boundaries and dislocations. Such effects become more pronounced at low temperatures, and may generally be detected, or even quantitatively measured, either by autoradiographic techniques or from the shape of the diffusion-penetration curve. Moreover, Hart (1957, R3) recently has shown that a considerable contribution of this nature may, at low temperature, be masked in what appears to be normal bulk diffusion. For these reasons, measurements of bulk diffusion should generally be limited to well-annealed single crystals and to temperatures near the melting point.

Careful attention must be paid to the chemical purity of diffusion specimens. Diffusion coefficients are quite sensitive to composition, especially in ionic compounds and semi-conductors, where minute concentrations of some kinds of impurities can have enormous effects on the diffusivity. Even in metallic systems, impurity levels of the order of 0.1 atomic percent can produce easily observable effects.

Temperature measurement and control must be performed with high accuracy. In typical situations, a temperature uncertainty of 1°C results in an uncertainty of several percent in the diffusion coefficient.

Of the various methods for determining the post-diffusion tracer profile, the sectioning technique is by far the most accurate and reliable, and should always be used where possible. However, because sectioning and counting of individual sections is a somewhat tedious procedure, many experimenters resort to the use of "simpler," indirect methods, such as surface counting, autoradiography, etc. It is now generally recognized that these indirect methods are subject to many possible sources of error (surface effects, back-scattering, absorption, etc.) which are difficult to evaluate quantitatively (Smith, Jaumot, and Arne, 1954, R3). The use of short-cut methods generally is made at the expense of accuracy.

In all cases, it is necessary to investigate carefully the radiochemical purity of the tracer material. When radioisotopes of a foreign element are present (either accidentally or as a daughter product), some method must be devised to eliminate or estimate their contribution to the measured activity. Even minute amounts of a fast-diffusing impurity may seriously alter the measured post-diffusion profile.

With proper attention to details, it has been possible to measure bulk tracer diffusion coefficients with a random statistical error

of a few percent. However, from an analysis of duplicate results reported by various investigators, it would appear that systematic errors, even in the best of measurements, probably limit the absolute accuracy of measured diffusivities to about ten percent. The corresponding uncertainties in the derived values of the activation energy, Q , and the frequency factor, D_0 , amount to about two percent and fifty percent, respectively. It should be emphasized that results of this accuracy are not obtained in a routine fashion, but require a high degree of experimental competence and patience.

Techniques for measuring diffusion of tracer atoms along grain boundaries are in general similar to those described for bulk diffusion, but with the following significant differences: 1) diffusion must be carried out at rather low temperatures in order to enhance the relative grain boundary contribution; 2) bicrystal specimens must be used if the information is to relate to the mobility of the tracer along a boundary of specified orientation or structure. Accurate absolute values of grain boundary diffusion coefficients are not obtainable, since the only measured parameter is the product of a diffusion coefficient and a geometrical factor (the "thickness" of the boundary) that can only be roughly estimated. However, since the geometrical factor is probably not strongly temperature dependent, it is still possible to obtain reasonably good values for the activation energy for grain boundary diffusion.

Precise methods have not yet been developed for measuring surface diffusion rates, although these are sorely needed. Difficulties arise in preparing reproducible surfaces, and in inhibiting direct vapor-phase transport of tracer atoms. The preparation and maintenance of impurity-free surfaces are also essential if true surface diffusivities are to be obtained. The recent development of super high vacuum techniques may prove a great aid in this connection. Surface diffusivities have thus far been measured in only a few systems, most notably by electrical methods in W-Th (Brattain and Becker, 1933, R1), and by tracer methods in silver (Nickerson and Parker, 1950, Winegard, 1953, R1). The field emission microscope also shows promise of being an extremely useful tool for measurement of surface migration rates in a limited number of materials (Hackerman and Simpson, 1956, R1).

2.2. Electrolytic Conductivity

In ionic solids and other insulating materials, mass transport has long been studied by measurement of the electrical conductivity and transference numbers. The results provide direct information

regarding the mobility of a particular charged species and hence have contributed greatly to the elucidation of the mechanism of transport processes in such materials. In fact, the very existence of electrolytic conduction constitutes direct evidence of crystalline disorder, and it is largely for this reason that the basic nature of this disorder and the role which lattice defects play in transport properties of ionic compounds were understood much earlier than in metals.

Techniques for measuring electrical conductivity (or resistivity) are well-known and highly precise so long as A.C. methods are employed. But the net mass flow that is required for the measurement of small transference numbers very often leads to many experimental problems (filamentary deposits, side-reactions, etc.), whose solution requires an extraordinary amount of experimental ingenuity. Since the techniques employed generally differ, depending on the material, it is difficult to make any generalizations in regard to the reliability of published data derived from conductivity and transference number measurements. However, in several of the alkali halides in particular, it has been fairly well established that the transference number of one of the constituent ions is essentially unity, at least over a rather broad temperature range. It is in these materials that conductivity measurements have been most actively pursued, with results of the same order of precision as in tracer diffusion measurements. Careful control of purity, structural imperfections and temperature must, of course, be exercised, as noted previously.

2.3. Concentration and Mobility of Defects

Many attempts have been made to deduce values of the energy of defect formation, ΔH_f , from the temperature dependence of various physical properties such as thermal expansion, specific heat and electrical resistivity. It is generally assumed that, in the absence of defects, the high temperature properties would be accurately represented by an extrapolation of low temperature data, and that the difference between the measured and extrapolated quantities is proportional to the defect concentration. Such extrapolations, in general, do not have a sound theoretical basis, and may introduce serious errors in the derived results. However, with a better understanding of the high-temperature properties of crystals, especially of the contributions resulting from anharmonic effects, such methods could conceivably become of greater value.

In principle, the defect concentrations present in thermal equi-

librium can be determined far more accurately by simultaneous dilatometric and X-ray measurements of the thermal expansion coefficients. The difference between the microscopic and macroscopic densities gives the fractional concentration of defects directly, the only assumption being that the defect can be considered as a center of isotropic elastic distortion. Even for temperatures near the melting point, the defect concentrations are apparently quite small, of the order of a tenth of one percent in metals (Simmons and Balluffi, 1959) and hence extreme precision is required for accurate data.

Currently, the study of the properties of rapidly quenched materials appears to be the most reliable method for directly investigating the properties of defects in metals (Bauerle and Koehler, 1957, R3). The defect concentration characteristic of the initial temperature of the specimen before quenching is deduced most accurately from low temperature measurements of the electrical resistivity, although other properties, such as volume and energy changes, have also been used for this purpose. From the dependence of the extra quenched-in resistivity on quenching temperature, ΔH_f may be evaluated, while studies of the kinetics of recovery following quenching yield information concerning the mobility of the defects.

Although quenching experiments appear simple in principle, they are extremely tedious and by no means free from complications. Because of the sensitivity of the electrical measurements to impurities and to mechanical deformation, the samples must remain uncontaminated during annealing and quenching, and the quenching must be carried out in such a manner as to prevent any deformation. Very fast quench rates ($\sim 10^4$ degrees/sec.) are required to prevent an appreciable fraction of defects from annealing out during the quench. Moreover, the kinetics of the annealing process are very complex and apparently sensitive to the quench temperature and to the presence of extremely small amounts of impurities. Despite these difficulties, the derived values of ΔH_f appear to be reproducible to within about 10 percent. The activation energies for defect motion, as deduced from kinetic measurements, must be considered much less certain.

In those ionic crystals in which cation vacancies are the predominant defects, studies of the dependence of diffusion and ionic conductivity on impurity content have been very useful. For instance, the introduction of divalent impurity cations in a monovalent ionic salt is accompanied by the creation of an equal number of cation vacancies in order to preserve electrical neutrality. At

sufficiently low temperatures, where the concentration of thermally-produced vacancies is much smaller than the divalent impurity concentration, the temperature dependence of diffusion or conductivity corresponds only to the energy of motion of the vacancies. But if the impurity concentration is not too large, there will be some temperature above which the thermal vacancies outnumber those artificially introduced, and then the observed activation energy will correspond to the sum $\Delta H_i + \Delta H_m$. Thus, assuming that the impurities have no direct effect on ΔH_i or ΔH_m , measurements of the diffusion coefficient or ionic conductivity over an appropriate temperature range may be used to evaluate the separate energies.

Radiation damage experiments have also provided a fruitful means of studying the properties of defects in solids. These measurements are discussed elsewhere in this volume.

2.4. Internal Friction

The anelastic effects long known to be associated with the presence of interstitial solutes in body-centered cubic materials are now widely accepted as resulting from the stress-induced redistribution of the interstitial atoms. On the basis of the model developed by Snoek (1941, R4), the time constant for this redistribution may be related directly to the diffusion coefficient of the interstitial. The model is confirmed (Wert, 1950, R4) by the excellent agreement which has been obtained with conventional diffusion measurements. At sufficiently low concentrations, where the relaxation process can be described by a single time constant, the method yields results of high accuracy; with variations of technique, it can be used to measure relaxation times over a range of many orders of magnitude, so that activation energies can be determined with a high degree of precision. However, the relaxation peak may often be complicated as a result of interactions between solute atoms, and care must always be exercised both in the preparation of samples and in the interpretation of the results.

Although an analogous relaxation process cannot be observed in dilute substitutional solid solutions, an internal friction peak which appears to be associated with diffusional motion has been found in some concentrated (~ 20 atomic percent) alloys. On the assumption that the relaxation is associated with stress-induced changes in local order, Norwick (1952, R3) has derived a relation between the relaxation time and the self-diffusion coefficients of the two constituents. Since tracer measurements are usually carried out at much higher temperatures than relaxation

measurements, a considerable extrapolation has been necessary to test the validity of this correlation. A rigorous test must await the careful measurement of all of the pertinent quantities in the same temperature range. Nevertheless, the method is particularly useful with respect to the small distances (or short times) over which diffusion is measureable, and is therefore of great value in investigations of diffusion at low temperatures and under nonequilibrium conditions.

2.5. Special Techniques

Recently a number of novel techniques have been introduced which show varying degrees of promise as methods for measuring diffusion rates in special cases. Slichter and co-workers (1952, R3) have employed nuclear magnetic resonance techniques to determine self-diffusion rates in a number of metals, specifically by measuring the changes in spin-lattice relaxation time and line width resulting from the thermal motion of atoms. Rathenau and associates (1957, R3) have determined diffusion rates for interstitial solutes in ferromagnetic domain walls. Sippel (1959) has determined diffusion rates in copper-gold alloys at very low temperatures by studying the dependence of the elastic scattering of protons and deuterons on the extent of diffusion of gold, obtaining results in good agreement with sectioning measurements.

Of these various techniques, nuclear magnetic resonance (NMR) appears to be most generally applicable to the study of diffusion in a large class of materials, particularly those which are not amenable to investigation by ordinary tracer methods. Since NMR measurements are concerned with the diffusional motion of atoms in the bulk of the lattice, they are relatively insensitive to effects occurring at grain boundaries or dislocations. NMR measurements can be made in the complete absence of chemical concentration gradients, and hence the results obtained are directly related to the true self-diffusion coefficient. In addition, since the relaxation resulting from single atomic jumps can be determined from NMR measurements, the technique permits determination of the actual jump frequencies of the diffusing atoms, correlated or not, whereas measurements involving actual mass flow yield only the rate of uncorrelated jumps. At present, while the absolute values of the diffusivities calculated from NMR measurements are subject to some uncertainty, because of limitations in the theoretical model, these methods nevertheless are capable of yielding quite precise values for the ratios of diffusivities of like substances.

Any solid state transformation whose rate is rigorously diffu-

sion-controlled can in principle serve as a basis for measuring diffusion rates. The evaluation of accurate diffusion coefficients, however, requires both a complete knowledge of the initial and time-dependent boundary conditions as well as a mathematical solution of Fick's Laws appropriate to these conditions. Rarely, if ever, are both of these factors known sufficiently well to place much reliance on the absolute magnitudes of the diffusion coefficients obtained in this manner. On the contrary, it is generally necessary in the analysis of kinetic data to make certain simplifying assumptions concerning either the boundary conditions or the solution of the differential equations or both, and diffusion data thus obtained are compared with known data as a check on the validity of these assumptions.

However, when the kinetics of a process may be described by a rate constant that is proportional to a diffusion coefficient over a reasonable range of temperatures, the activation energy for the overall process may be taken to be that for diffusion. As an example, activation energies for high-temperature creep and sintering of pure metals are frequently found to be in good agreement with corresponding activation energies for self-diffusion derived from tracer experiments. Although these indirect methods may be of some value in making order-of-magnitude estimates of data not otherwise available, they have not yet reached the stage of development where they may properly be considered as experimental techniques for acquiring diffusion data.

3. ATOMIC MODELS

In this section we are concerned primarily with those experiments, concepts, and theories which most directly bear on the atomistic nature of diffusion processes, particularly bulk or volume diffusion processes in single crystals. One of the principal aims of basic diffusion studies has been to formulate a detailed atomistic model or mechanism for the unit jump process. The mechanisms which have been proposed for diffusion in solids of various structural types will now be presented and the supporting evidence reviewed.

3.1. Reaction Rate Problem for Diffusion in Solids

Irrespective of the detailed mechanism, all atomistic jumps may be described within a common conceptual framework. The diffusion coefficient is given by

$$D = \alpha r^2 \nu f \quad (3.1)$$

where r is the effective jump distance, α is a constant determined by the geometry of the lattice, and ν is the jump frequency. The factor f represents a correction which arises from the fact that successive atom jumps may be correlated. Each diffusion jump starts with the atoms at their equilibrium positions. A lattice defect may or may not be present in the vicinity. The atoms then pass through an intervening configuration of higher energy, corresponding to a saddle point in phase space, and finally relax into a new equilibrium configuration. It would appear that standard reaction rate theory could be applied to those processes; however, we are not dealing with a large assemblage of small independent systems, each of definite size and composition. The "equilibrium system" considered in any particular type of jump in the solid may involve a large number of atoms, the number depending upon the lifetime of the saddle-point configuration.

One procedure for handling the reaction rate problem for diffusion in solids has been to take the description of the activated complexes afforded by statistical mechanics and to ascribe thermodynamic functions to these complexes, as though each were a small sub-system in equilibrium with the lattice. It has become customary to evaluate the intrinsic changes in the free energy, G , the enthalpy, H , the entropy, S , and the volume, V , that are associated with the formation of the activated state. The jump frequency is then given by

$$\nu = K p_d \nu_0 e^{-\Delta G_M^*/RT} \quad (3.2)$$

where p_d is the probability that a lattice defect is present, if required by the mechanism, ν_0 is a weighted mean frequency (Vineyard, 1957, R3), and ΔG_M^* is the increase in Gibbs free energy of the system for the transition between the equilibrium and saddle-point configurations (the asterisk indicates that the degree of freedom associated with the jump direction is not included). The factor K represents the transmission coefficient, i.e., the probability that an atom which has sufficient energy to surmount the barrier will, in fact, execute a complete jump. Generally, K is considered unity, without much justification, but if the lifetime of the excited state is fairly short, the transmission coefficient may be appreciably less than unity.

The change in free energy, ΔG_M^* , may be separated into enthalpy and entropy terms using the familiar relation

$$\Delta G_M^* = \Delta H_M^* - T \Delta S_M^* \quad (3.3)$$

It can be shown that $-R \left[\frac{\partial (\ln \nu / p_d)}{\partial (1/T)} \right]_P$ is ΔH_M^* , the energy needed

to bring the system from the equilibrium to the saddle point configuration under isothermal isobaric conditions. On the other hand, ΔS_M^* may be evaluated, in principle, from a knowledge of how the vibrational spectrum of the crystal is altered on going from the equilibrium to the saddle-point configuration. At high temperature, ΔS_M^* is calculated from the expression

$$\Delta S_M^* = k \sum_i \ln (v_{ie}/v_{is}), \quad (3.4)$$

where the indices e and s refer to equilibrium and saddle point respectively; k is the Boltzmann constant, and the index i numbers the vibrational frequencies of the lattice. In regions near the defect, an increase of atomic crowding in the activated complex causes an increase in the vibrational frequencies, giving rise to a decrease in ΔS_M^* and consequently in D_0 . The strains at large distances from a point defect are purely elastic and the corresponding entropy change is always positive. An effort has been made to estimate the intrinsic entropies of several proposed saddle-point configurations for diffusion (including ΔS_F , the entropy for defect formation) and, although the calculated entropies are mostly positive, the values obtained are systematically too low to give agreement with the experimentally observed D_0 (Huntington, Shirn and Wajda, 1955, R3). While the accuracy of these calculations is uncertain (mainly because of drastic assumptions made in treating the vibrational spectrum), it is possible that the difficulty arises from a fundamental inadequacy of the model.

Another line of reasoning based on the thermodynamic relation

$$dG = -SdT + VdP \quad (3.5)$$

leads to the expression

$$\Delta V_M = -RT \left[\frac{\partial (\ln v/P_d)}{\partial P} \right]_T. \quad (3.6)$$

Here ΔV_M is the increase in the volume of the system at the saddle point relative to that of the equilibrium configuration. It would be expected that ΔV_M would ordinarily be positive, because of the strong repulsive forces between a diffusing atom and its neighbors, but the experimental evidence, to which we shall return in a later section, gives rather small values for the total activation volume, ΔV . The latter also includes the volume of formation of the defect, ΔV_F , and is determined from the pressure dependence of the diffusion coefficient, as

$$\Delta V = \Delta V_M + \Delta V_F = -RT \left(\frac{\partial \ln D}{\partial P} \right)_T. \quad (3.7)$$

It is apparently not necessary for the region in equilibrium with the saddle-point configuration to extend to the surface to validate Eq. (3.6). This expression will still be valid if the excited state has a lifetime such that the dilation extends only to the Hooke's Law region in the crystal, since the dilation vanishes at larger distances. (Of course, the ΔV_M in this case would not include the increase in volume arising from the surface boundary condition.)

The basic assumption made in the preceding treatment is that the diffusing atom is in thermal equilibrium with the lattice during the elementary jump, and hence that the lifetime of the excited state is sufficiently long to ascribe thermodynamic parameters to the saddle-point configuration. Diffusion measurements do not give direct information regarding the lifetime of the excited state. However, there is some evidence, based on the line width of gamma rays emitted from short-lived isomers in solid lattices, which suggests that the lifetime of the excited state may be less than 10^{-12} seconds. If this is true, a serious question may arise regarding the validity of the assumption of thermal equilibrium.

Recently, Rice (1959, R3) has proposed a different approach to the reaction rate problem, based on consideration of the dynamics of the diffusing atom. In this treatment, the motions of the diffusing atom and other atoms most closely involved in the jump are resolved into their components in terms of the normal modes of vibration of the lattice. Considering an harmonic crystal, and making somewhat arbitrary assumptions about the irreversibility of the process, Rice finds that the jump frequency varies exponentially with temperature; however, the dominant terms in his expression for the activation energy for motion cannot be identified with changes in Gibbs free energy, unlike the result of the thermodynamic equilibrium treatment. The application of this dynamical approach to real crystals appears formidable, but the method may provide considerable insight, particularly in delineating means for refining the thermodynamic equilibrium treatment to overcome the possible difficulties discussed above. What is needed is a return to a more fundamental kinetic treatment to handle the activation process.

In general, a combination of both statistical mechanics and kinetic theory is needed to treat these complex reaction-rate problems. Statistical mechanics gives the probability of formation of the activated complex, but it is necessary to resort to kinetic theory in order to determine whether any particular configuration will correspond to a traverse of the saddle point or to a reflection.

Evidently, the reaction rate problem is still not satisfactorily resolved.

3.2. Diffusion Mechanisms

There are three atomistic mechanisms which have traditionally received primary consideration: direct interchange, interstitial motion, and vacancy migration. However, more complex models have frequently been invoked. Direct interchange is actually a special and improbable case of ring diffusion for which $n = 2$. Ring mechanisms for which $n = 4$ have received most attention (Zener, 1959, R3), although in some materials the possibility exists that the ring rotation may involve only three atoms. Ring mechanisms are the only ones which do not require the prior presence of a lattice defect in a nearest-neighbor position in order for an atom jump to occur.

The interstitial mechanism in metals is limited primarily to the diffusion of small impurity atoms which dissolve interstitially. This is because the energy required for the formation of an interstitial is generally quite large for atoms which have approximately the same size as the solvent. Appreciable concentrations of interstitials may be introduced in pure metals, however, by irradiation or by cold work. Under these conditions, the solvent atom may migrate as an "interstitialcy," in which a nearest neighbor of the interstitial becomes displaced to a new interstitial position, while the old interstitial relaxes to a normal lattice position. There is also evidence for interstitialcy motion in the transport of the cation in silver salts. Another possible type of interstitial is the crowdion (Paneth, 1950, R3), which consists of an extra atom inserted into a line of atoms in a close-packed direction. Recent developments indicate that the crowdion may be of importance in the annealing of radiation-damaged metals.

The following assignments for the diffusion mechanism are generally accepted for some common types of solids. In ionic crystals, only those defects which preserve charge neutrality for the crystal as a whole are allowed. The Schottky defect (separate positive and negative ion vacancies) is preferred for most alkali halides, but the Frenkel defect (positive ion interstitial and vacancy) is favored for diffusion in the silver salts. In both cases, it is the positive ion which is usually the more mobile. The fact that substitutional diffusion is generally slow in semiconductors suggests that a vacancy mechanism is operative. On the other hand, certain elements, such as copper and lithium, can diffuse with great rapidity in elemental semiconductors, apparently by

moving on an interstitial sublattice. One manifestation of this phenomenon is that the rate of precipitation of copper in germanium at low temperatures is limited by the disappearance of residual vacancies which tend to trap the fast-moving interstitials.

For metals and alloys the vacancy mechanism is largely preferred except for interstitial impurities. However, recent diffusion measurements in body-centered cubic alloys (Paxton and Kunitake, to be published; Paxton and Pasierb, 1960), as well as some results noted in sintering experiments in iron, do not appear to be consistent with a vacancy mechanism, and have been interpreted in terms of a ring mechanism. Considerable evidence exists for the presence of both positive and negative ion vacancies in oxides and spinels. However, the accuracy of experimental measurements of diffusion in these materials is hampered by the lack of availability of good single crystals as well as by small departures from stoichiometry, and it is generally not possible to assign specific diffusion mechanisms on the basis of current data.

It must be stressed that the assignment of a particular diffusion mechanism is rarely a straightforward matter. Information concerning the mechanism and the associated defect structure is generally inferred from a study of many properties by a wide variety of experimental techniques, no one of which permits a unique assignment. Because of the indirect character of most of the evidence, it seems reasonable to urge workers in this field to maintain a persistently critical point of view.

3.3. Temperature and Pressure Dependence of the Diffusivity

For bulk diffusion, it has been found experimentally in nearly every material studied that ΔH^* , given by

$$\Delta H^* = -R \left(\frac{\partial (\ln D)}{\partial (1/T)} \right)_P, \quad (3.8)$$

is constant over a wide range of temperature. Perhaps the most striking example is the diffusion of carbon in iron for which a linear dependence of $\ln D$ vs T^{-1} is obtained over twelve decades of D . It follows that in such a situation ΔS^* is also temperature independent. From the data available, it is tempting to conclude that the activation enthalpies for diffusion in solids are, within experimental error, determined only by lattice structure and binding forces, and are independent of thermal vibrations and thermal expansion within the temperature range over which measurements are usually made.

The pressure dependence of the diffusivity has been studied in a number of systems, including white phosphorus, sodium, lead, silver bromide, and some Ag-Zn alloys. In each case, the diffusivity is found to decrease exponentially with increasing pressure. The total activation volume, $\Delta V_F + \Delta V_M$ of Eq. (3.7), falls in a range from 0.4 to 0.7 of the respective atomic volume, which seems somewhat small. For those liquid metals that have been studied, the total activation volume is about 0.05 atomic volumes. Measurements on interstitial diffusion of nitrogen and carbon in iron have indicated that ΔV_M is very small (~ 0.01 atomic volumes) for these cases (Bosman, Brommer and Rathenau, 1957, R3). However, larger values of ΔV_M (~ 0.15 atomic volumes) have been found for diffusion of interstitial oxygen in vanadium (Tichelaar and Lazarus, 1959, R3). The ΔV_F for either vacancy or interstitial diffusion would be expected to be positive for close-packed crystals, although not necessarily so for more open structures. It would seem that ΔV_M should also be positive but its magnitude is apparently quite small.

3.4. Theoretical Approaches to Atomistic Mechanisms

Theoretical attempts to make quantitative predictions about atom movements should be viewed critically. Such attempts have been most useful for the alkali halides, where a simple and reasonably valid model for the perfect lattice exists. It has therefore been possible to calculate fairly rigorously the energy of formation of Schottky defects, the energy corresponding to saddle-point configurations, and the energy of association between positive ion vacancies and divalent impurities. Enough is known about the alkali metals so that defects of high symmetry could be handled, albeit with considerable labor, to give reasonably reliable results. With several other metals, some general conclusions can be drawn on the basis of semi-empirical, two-body force laws. In general, these considerations indicate that the vacancy mechanism should require less energy than either an interstitial or ring rotation mechanism. The calculation also gives some estimates of the change in volume to be expected for different lattice defects. The following limitations of the theory should be borne in mind:

- (1) Only a very few materials have been analyzed theoretically with any high degree of completeness, and then always by somewhat oversimplified models. Even for the alkali halides, complex many-body forces must be introduced to explain the departure of the elastic constants from the Cauchy relations, and these have not been considered in the calculations.

(2) The more complex models, involving crystal symmetry and electronic band structure, become unworkable when applied to defect configurations.

(3) The atomic configurations of the most interest for diffusion, such as the activated complexes, possess a low degree of symmetry which makes the application of even simple models laborious.

(4) Uncertainty is introduced because semi-empirical force laws must be extrapolated considerably beyond the range of interaction for which they were developed.

(5) In the neighborhood of a defect in metals, the interaction between the distorted ionic lattice and the modulated electronic distribution seriously complicates the attainment of a mutually self-consistent arrangement.

Because of these limitations in a theory which tries to treat the binding forces in a solid in a careful and fundamental way, it appears worthwhile to explore the predictions arrived at with cruder models such as those based on only two-body central forces, augmented perhaps by a volume force (internal negative pressure). There are many situations where simplified models of this type may provide valuable insight, but the investigator must always recognize the approximate nature of his model. As the interactions between lattice defects assume greater and greater importance, quasi-theoretical methods will be needed to explore complicated atomic configurations, e.g., diffusion along dislocations.

3.5. Influence of Chemical, Electrical and Thermal Gradients

The motion of markers within the diffusion zone in the presence of chemical gradients (Kirkendall effect) is the single clearest piece of evidence that we have concerning the nature of atom movements. In the systems where such motion has been detected, which include every class of materials so far investigated—ionic salts and oxides as well as metals, it is clear that ring mechanisms cannot provide the sole contribution to diffusion. In most systems that have been studied, macroscopic voids have also been observed on the side of the original diffusion interface which contains the more mobile constituent. The presence of voids provides some evidence for the vacancy mechanism, since these voids may reasonably be supposed to result from a supersaturation of the defect responsible for diffusion on the annihilation side of the interface. The vacancy model requires that the net flow of atoms from the side of the interface containing the more mobile constituent be counterbalanced by a net flow of vacancies into this region. It has

been found that moderate external pressure will inhibit the development of porosity (Barnes and Mazey, 1958, R3). Moreover, it has been shown that dislocations may act as sources for vacancies and also as sinks for those vacancies which do not aggregate or condense to form voids, but the details of these processes are not well understood.

In materials such as the alkali halides, where electronic conductivity can be neglected, much has been learned concerning the nature of ionic transport from studies of the electrical conductivity as a function of temperature, purity and thermal history. By a judicious use of the Nernst-Einstein relation, the measured conductivity can be related to the diffusivity. Deviations from this relation can be expected if any mechanism exists which may contribute to the transport of matter without transport of charge, or vice versa. Thus, if the association between positive- and negative-ion vacancies or between divalent impurities and vacancies leads to the formation of a mobile neutral complex, the observed diffusion coefficient will be greater than that calculated from the conductivity data. Deviations of the opposite sign will occur whenever the process of atom jumping moves the effective charge through a mean-square distance which is greater than the mean-square displacements of the atoms involved in the jump. This situation is illustrated by interstitialcy diffusion. The correlation between successive atomic jumps also acts to lower D below the value predicted from the Einstein relation.

In order to specify the diffusion mechanism, the fraction of the total charge which is transported by the various possible carriers must be known. The technique of measuring transport numbers has been applied most successfully to the alkali halides, in general clearly indicating that the major part of the current is carried by the positive ion, except near the melting point where both ions appear to participate. Positive ion mobility, plus optical evidence for the existence of negative ion vacancies in the alkali halides, points to the presence of Schottky defects. In the silver salts, it has been observed that only the silver ion contributes to the conductivity even at high temperatures, which indicates the existence of Frenkel defects for the positive ions rather than Schottky defects, since the latter would require the presence of negative-ion vacancies.

The study of mass movement under the influence of an electric field has been pursued most strongly in Europe. Like the electrolysis of ionic salts, only to a much smaller degree, the passage of electric current through alloys leads to some separation of con-

stituents. It has been customary to speak of transport numbers in connection with this motion, and to attribute definite charge valences to impurity atoms, such as carbon in iron. An interesting recent development in this field has been the experiment of Seith and Wever (1958, R4), who found, by studying the displacement of phase boundaries and markers, that a mass motion of both constituents occurs in Cu-Al alloys when subjected to the passage of high electric current. In the β -phase, which is an electron conductor, the motion is toward the anode, and in the α -phase, which is a hole conductor, the motion is reversed. Seith and Wever have postulated that momentum exchange with the drift of charge carriers may be responsible for this behavior. Experiments on gold wires have shown motion toward the anode proportional to the current density, and having approximately the same temperature dependence as the self-diffusion coefficient. This technique seems to offer considerable promise as a means of learning something about the specific resistivity of the activated complex for diffusion.

Mass transport induced by a thermal gradient is called the Soret effect. This phenomenon occurs in gases, liquids and solids. Since the effect depends on the character of the thermal contact between the high and low temperature regions, it cannot be explained in terms of equilibrium parameters. Instead, it is customary to describe the Soret effect in terms of the quantity Q_i^* , called the heat of transfer, and defined as the heat that must be injected to maintain a steady state at any point from which a particle of the i th constituent is removed. According to the Second Law of Thermodynamics, constituents with $Q^* < 0$ move to the hot end and those with positive Q^* to the cold end of the system. The quantity Q^* can be resolved into an equilibrium term, the specific enthalpy per particle of the diffusing species, and a transport term which depends on the nature of the jump process.

Interest centers on the Soret effect for interstitial alloys, substitutional alloys, and for vacancy flow. For interstitial diffusion, Q^* has been found to be negative for carbon and nitrogen in iron, but positive for the motion of hydrogen in α -zirconium. The behavior of hydrogen in zirconium can be understood in terms of a model which assumes that the hydrogen atom carries the activation energy of motion with it to the new position, with little loss of energy to the lattice. For larger interstitial atoms, such as carbon and nitrogen in iron, the activation energy is clearly shared with lattice atoms near the path of the jump. From a kinetic standpoint, it can be imagined that the interstitials move in the

direction in which favorable configurations for jumping are created more frequently because more thermal energy is available. Both positive and negative heats of transfer have been found in substitutional zinc alloys. Several attempts have been made to demonstrate a Soret effect for vacancies, using surface markers on specimens in a strong thermal gradient, but to date no such effect has been observed.

3.6. Diffusion of Impurities

Although early evidence indicated that the diffusion of solutes or impurities in dilute solid solution was much more rapid than self-diffusion, several experiments carried out in recent years using radioactive isotopes have shown that the differences in activation energies between impurity and self-diffusion are much less than were formerly supposed (Tomizuka and Slifkin, 1954, R3), and also that these differences may be positive as well as negative. Activation energies greater than that for self-diffusion in the noble metals have been observed for solutes which belong to a column of the periodic table lying to the left of the solvent.

For the diffusion of various solutes in silver, the activation energies associated with elements lying to the right of silver in the same row of the periodic table show a monotonic decrease with increasing valence. Lazarus (R3) has shown that this behavior agrees well with what would be expected for the sum of ΔH_F and ΔH_M on the basis of a simple model which regards the impurity as a screened point charge. The motion itself is treated as a simple shearing action. Although the treatment is quite approximate throughout, subsequent workers have concentrated on refining the choice of the screening parameter, originally based on the Fermi-Thomas model, usually to the detriment of the agreement with experiment. The simple model has given less satisfactory agreement with experiment when applied to diffusion in copper-base alloys, in which a more erratic variation of activation energy with valence is observed. For the transition elements lying to the left of copper in the periodic table, the activation energies are all higher than that for self-diffusion, but increase in order on going from iron to nickel—exactly opposite to what would be expected if the relative position in the periodic table were the only determining factor. The uncertainties in assigning a definite valence to the transition elements have long been recognized; nevertheless, it must be concluded from the general behavior noted above that other factors aside from valence also play a role in determining diffusion rates.

The "size" of the impurity atom is a natural choice for such a factor. Swalin (1957, R3) has considered how impurity size should affect the energy for vacancy formation and motion to a neighboring site, and finds qualitative agreement for the diffusion of various elements in nickel.

More detailed treatments of impurity diffusion are made difficult by uncertainty regarding the electronic state of the solute atom and its surroundings. The observed differences between the diffusivities of various impurities in a common solvent are generally small, and experimental errors may be obscuring some of the detail, making comparison between theory and experiment quite difficult.

Solute diffusion that is much more rapid than self-diffusion cannot be explained solely by a more ready exchange between foreign atoms and vacancies. It is necessary to postulate in addition that a binding energy exists between a vacancy and solute atom. For strong binding, the vacancy and solute atom would diffuse together as a unit or molecule (Johnson's mechanism), but it appears more likely, on the basis of current data, that in most situations the partnership may be short lived. On the other hand, correlation effects also become much more important when considering the diffusion of solute or impurity atoms, and this may serve to reduce the role of diffusion of impurities, since the vacancy interchanges more frequently with the impurity atom than with their mutual neighbors. The precise manner in which the relative frequency of these interchanges may be affected by valence and size considerations represents a challenging theoretical problem.

Because correlation effects differ depending upon the diffusion mechanism, it is possible, at least in principle, to utilize these differences to distinguish between various possible mechanisms. It has been pointed out by Schoen (1959) that from a comparison of the diffusion rates of two radioisotopes of the same element, one can obtain an upper limit to the correlation factor. Schoen has applied this argument to show that the diffusion of cadmium in silver and in copper cannot take place by an interstitial mechanism, for which the correlation factor would be unity. The approach appears quite promising particularly if future developments in reaction rate theory show that the upper limit determined by diffusion experiments using isotopes which differ only in mass lies close to the actual correlation factor.

A study of the diffusion of substitutional impurities in hexagonal and body-centered cubic materials should prove most en-

lightening. The body-centered cubic structure is particularly interesting, since in the undistorted lattice, the impurity-vacancy complex could not move as a unit by single vacancy jumps; the impurity and vacancy must dissociate if the motion is to be decorrelated.

Impurities also influence the self-diffusion coefficient of the solvent, usually toward larger values, as shown by the work of Hoffman, Turnbull, and Hart (1955, R3) on dilute silver-base alloys. Increased solute diffusivity by the vacancy mechanism implies greater mobility of solvent atoms as well. Solute mobility also increases with solute concentration, but not as rapidly as for the solvent. Even those impurities which do not enhance the vacancy concentration might be expected to increase solvent mobility by the action of their stress fields, according to Overhauser (1953, R3). However, it is known that at least one impurity, namely palladium, reduces the self-diffusion coefficient of silver (Nachtrieb, Petit, and Wehrenberg, 1957, R3).

Self-diffusion in elemental semiconductors is enhanced by doping with donor impurities and reduced by the introduction of acceptors. This effect argues strongly for the vacancy mechanism, since it is known that the vacancy functions as an acceptor. When the Fermi level is raised by the addition of an impurity, the energy needed to form the vacancy is decreased. Self-diffusion and substitutional diffusion in these materials, and also in the III-V compounds, are characterized by large activation energies and large D_0 values. Presumably, the formation and motion of vacancies are accompanied by rather large changes in the energy and entropy which arise from the breaking of bonds. In III-V compounds it would appear that each constituent may diffuse separately on its own sublattice. On the other hand, small atoms like lithium diffuse with great rapidity, presumably by an interstitial mechanism, and other elements, particularly copper, act as slow-moving substitutional impurities part of the time and as highly mobile interstitials for the remainder.

3.7. Diffusion in Alloys

In concentrated alloys, it is no longer meaningful to consider the association of a vacancy with a particular solute atom. The tightly-bound complex characteristic of a pure Johnson mechanism would cause a flow of vacancies and solute atoms in the same direction, whereas the existence of the Kirkendall effect implies a counterflow of vacancies and atoms of the more mobile constituent.

Where measurements have been made separately of the energies

of formation and motion of vacancies, for example by quenching techniques, it appears that the two are nearly equal for pure metals, but that the energy of motion is about twice the formation energy in concentrated alloys.

Some interesting features arise for diffusion in ordered alloys. The importance of a defect structure in enhancing diffusion was first demonstrated in the Al-Ni system (Nix and Jaumot, 1951, R4). In β -brass below the ordering temperature, it has been found that zinc diffuses more rapidly than copper (Kuper et al., 1956, R1). At first, this appeared to preclude the possibility of nearest-neighbor vacancy jumping, but it was pointed out later that by taking into account the departure from complete order at the temperatures involved in the experiment, the data could be reconciled with a vacancy mechanism. Single jumps between nearest-neighbor positions do not maintain the energy of the system constant, since a local disordering results. However, it may be postulated that the elementary jump process involves a sequence of single jumps which alternately order and disorder the lattice. As these sequences differ slightly according to which constituent is diffusing, there enters another possibility for explaining a difference in diffusion rates up to a factor of two for an NaCl structure, or a factor of 1.5 for a CsCl lattice, while still retaining the nearest-neighbor vacancy mechanism (Elcock and McCombie, 1958, R4).

3.8. Anisotropic Diffusion

A study of the anisotropy of diffusion in non-cubic materials offers some possibility that the mechanism can be identified on purely geometrical grounds. To date, a clear indication of the mechanism has been forthcoming only with zinc, in which a marked difference in the activation energies perpendicular and parallel to the hexagonal axis has been observed. The evidence points strongly towards the vacancy mechanism. In other close-packed hexagonal metals (Mg, Ti), the activation energy has proved to be nearly isotropic. Data obtained on other non-cubic metals (Bi, Sb, and In), have so far given little definite indication of the mechanism. For In, which is tetragonal, the two diffusion coefficients differ by only about 30% throughout the diffusion range, while the activation energies vary markedly with direction in both Bi and Sb. The anisotropic aspects of diffusion in high melting point metals (U, Co) have not been investigated and anisotropic diffusion of impurities has been almost completely neglected.

4. PHENOMENOLOGICAL AND SEMI-EMPIRICAL MODELS

The purpose of this section is to examine some of the approaches which have been made in developing phenomenological and semi-empirical models, with the intent of pointing out their general applicability and limitations. Most such models or correlations have been developed for and applied to metals, but many have also been extended to semiconductors and ionic compounds as well.

The correlations seem to fall into two categories: (a) those derived without any consideration of mechanism, and (b) those derived on the basis of an assumed mechanism. Not infrequently, mechanism has been inferred on the basis of a purely empirical relation of type (a), but this approach has little validity. Correlations of this type must be considered useful only for estimating the diffusion parameters for an unstudied system. Such correlations are usually as good and sometimes better for this purpose than are the more detailed models of type (b). Correlations of the second type have played a significant role in elucidating the diffusion mechanism and in establishing the accuracy of experimental data, although it must be pointed out that the predictions are usually inferential, since an empirical proportionality constant generally appears somewhere in the derivation and this constant can only be evaluated experimentally. In the scientific literature, these semi-empirical theories have frequently been abused by investigators who, in attempting to understand their experimental data, have applied the theories indiscriminately, forgetting the underlying assumptions on which the equations were derived. It seems a questionable procedure to expect a theory of this type to predict the various diffusion parameters with a high degree of accuracy. Furthermore, the validity of experimental results should not be judged by their consistency or inconsistency relative to a given theory.

4.1. Correlation Between the Diffusion Coefficient and the Melting Point

An examination of the diffusion literature indicates that, for solids having the same crystal structure, the diffusion coefficients are frequently about equal at the respective melting points, although there are significant exceptions to this rule. For example, the self-diffusion coefficients for most face-centered cubic metals are generally about 3×10^{-9} cm²/sec at their melting points, although the diffusion coefficients for lead and gold are about an order of magnitude lower and an order of magnitude higher, respectively. A close correlation between the self-diffusion coeffi-

cient and the melting point has also been observed in the silver-palladium system, in which the diffusivity of silver is found to be essentially constant along the solidus line up to 22 percent palladium. A similar trend holds for the gold-nickel system, but not for silver-cadmium and silver-indium.

A generalization has frequently been made for cation diffusion in oxides that, at the melting point, $D \cong 10^{-7}$ cm²/sec. From the experimental data available, this relation seems to be valid within an order of magnitude. No inference is justified about possible mechanisms, except to suggest that cation diffusion might occur by the same mechanism in the various oxides studied.

4.2. Correlations Between the Activation Energy and Various Physical Properties of the Crystal

A large number of empirical correlations have been noted between the activation energy for diffusion and various physical properties of a crystal, such as the melting point, heat of fusion, heat of sublimation, surface energy, elastic modulus, etc. Some of these correlations predict Q with a considerable degree of accuracy. In all of these correlations, the activation energy and the physical property in question are related by a simple equation

$$Q = A \text{ (Property)} \quad (4.1)$$

where A is an appropriate proportionality constant. The general validity of such equations has been well established for self-diffusion in pure crystals, but the situation with regard to alloys is still open to question. Detailed considerations regarding diffusion mechanisms are hardly justified by the success of any of these correlations. The proportionality constants are strictly empirical, and have not yet been derived on the basis of an atomistic model. The only inference which can be drawn with any degree of confidence is that the activation energy for diffusion is directly related to the binding energy of the lattice, as are the energies for the other physical processes considered. The basic causal relationship connecting these observations remains obscure.

There is a certain danger in the use of such correlations, attributable perhaps to the complacency that sometimes arises when agreement is obtained with apparently well established criteria. As evidence, we can cite the numerous empirical correlations which were proposed before 1950 to rationalize solute diffusion data. We now know that the data on which these correlations were based were frequently inaccurate.

Two currently useful correlations are those relating Q for self-

diffusion with the melting point and the heat of sublimation for the pure material. For the relation between the activation energy and the absolute melting temperature, the constant A in Eq. (4.1) is 37 cal/mole-deg K. For the corresponding relation between Q and the heat of sublimation, A has the value 0.64. The correlation between Q and the melting point is somewhat more consistent than the correlation with the heat of sublimation, possibly due to the greater accuracy with which melting points have been measured for most materials.

More recently, a correlation between the activation energy and the heat of fusion has been suggested, with A equal to 16.5. Somewhat surprisingly, this correlation holds quite well for both face-centered cubic and body-centered cubic materials, as well as for the more complex structure of white phosphorus. Over a range of activation energies from 10 to 65 kcal/mole, the agreement is generally within ± 20 percent.

A correlation between the activation energy and the elastic modulus of the parent crystal is also expected, since there is a correlation between the modulus and melting point. This correlation has been carried considerably beyond the simple relationship of Eq. (4.1), and will be discussed in section 4.4.

4.3. Correlations Between the Diffusion Coefficient and the Activation Energy

One of the best known relations of this type is the Langmuir-Dushman equation, which is essentially a relation between $\log D$ and the activation energy. The Langmuir-Dushman equation is frequently used to estimate Q when the diffusion coefficient is known at only a single temperature, the success achieved being little more than a reflection of the fact that D_0 is usually equal to about 1. The validity of the expression for D_0 is therefore an accidental consequence of the parameters chosen. This equation should be looked upon as a strictly empirical relation. It is certainly groundless to attempt to calculate jump distances on the basis of the Langmuir-Dushman equation, as has been done by some investigators.

4.4. Correlations Based on the Elastic Modulus

4.4.1. Relation Between Q and Modulus

As noted, it is not surprising that a correlation between Q and the elastic modulus is found to exist. Because elastic behavior is better understood than melting, theoretical work has been carried much further in this general area, and, as a result, considerable

support for the vacancy mechanism of diffusion in metals has been obtained. A basic postulate of most theories of this type is that the barrier to diffusion can be considered as a geometric barrier between the diffusing atom and the adjacent vacancy, and that the jump energy can be equated to the strain energy needed to surmount this barrier. However, the consistency of these equations does not confirm the validity of the vacancy mechanism, since appeal must be made to diffusion experiments to evaluate certain parameters.

It appears that too much reliance is sometimes placed on the results of such calculations, ignoring some questionable assumptions which are made. For example, the elastic moduli used in the calculations are determined from macroscopic measurements on bulk crystals, but since diffusion presumably entails the motion of some type of defect, it is the force constants in the vicinity of the defect itself which are more relevant.

A further and more fundamental problem exists, stemming from the general assumption of Hookean behavior in the vicinity of the geometric barrier. This assumption is undoubtedly incorrect, since the local strains are much too large to be treated elastically. At least one attempt has been made to circumvent this problem by considering only those strains which are remote from the immediate vicinity of the diffusing atom to be elastic. This approach is unquestionably more valid.

4.4.2. Correlation Between D_0 and Q

One of the most significant advances in diffusion theory was provided by Zener (1952, R3) in his treatment of the relationship between D_0 and Q . The diffusing atom, in thermal equilibrium with the lattice, is assumed to surmount a free energy barrier which is due entirely to elastic strains. A correlation between D_0 and Q follows irrespective of mechanism. The assumption that the barrier depends only on elastic interactions is somewhat questionable, as are a number of other assumptions of the model. For example, the frequency term appearing in the expression for D_0 is probably not the Debye or Einstein frequency of the parent crystal, as is usually assumed. The vibrational frequency must at least be altered in the vicinity of the defect, but even in the absence of defects, the Debye or Einstein frequency is not obviously the best choice.

In spite of these difficulties, a strong correlation does seem to exist between D_0 and Q for self-diffusion in pure elements, as predicted by the theory. An exact comparison is difficult to make

because experimental values of D_0 are rarely known to within better than a factor of two. Consequently, it appears that detailed theorizing regarding the significance of experimental values of D_0 may be unwarranted. A good example of the problems which arise in interpretation is provided by solute diffusion studies. Current and seemingly valid sets of diffusion data for various impurities in silver (R3) are in marked disagreement as to the detailed correlation of D_0 with Q . Some data indicate an exponential proportionality as predicted, and others indicate that only D_0 varies with atomic size and other factors, while Q remains virtually invariant.

4.5. Correlations for Solute Diffusion

Interpretation of the activation energies for solute diffusion has centered mainly around application of the concepts of valence and the relative sizes of solute and solvent atoms. Both approaches rest on the assumption that the barrier to motion is entirely elastic, and hence are subject to the same limitations discussed above.

Several correlations relating solute diffusion coefficients to solute atom size have been suggested in the past, mainly on the basis of pre-1950 data, but most of these have since proved unreliable. The problem of the size effect is inherently quite complicated. It is expected that a wrong-sized solute atom in solution will produce a considerable distortion of the local environment which could in turn facilitate more rapid diffusion. On the other hand, if the impurity atom is larger than the solvent atoms, more difficulty might be experienced in surmounting the geometric barrier to motion.

For most systems, accurate determinations of the effective sizes of dissolved solute atoms are not available. The atom is also probably polarized upon passing through the barrier. The influence of this distortion is, of course, difficult to estimate with high accuracy. For a detailed treatment of the solute diffusion problem, more knowledge is needed concerning the theory of alloys in general.

A recent approach by Turnbull and Hoffman (1959) suggests a proportionality between the activation energy for solute diffusion and the binding energy of the solute in the solvent lattice as determined from thermodynamic measurements. An equation relating the diffusion coefficient to the heat of solution of the solute has been derived on the basis of quasi-chemical theory. The model correctly explains the sign of the observed effect, but provides

only fair quantitative agreement. The limitations of this approach are basically those inherent in the quasi-chemical theory, namely that only nearest-neighbor interactions are considered.

4.6. Correlations with Concentration and Effect of Concentration Gradients

4.6.1. Phenomenological Treatment Involving Concentration Gradients

It has been demonstrated experimentally that a chemical concentration gradient has a considerable influence on the diffusion coefficients of the various constituents (e.g., Manning, 1959, R3). Darken's equations relate the chemical diffusion coefficient to the self-diffusion coefficients obtained by radioactive tracer techniques. Darken's analysis rests on the following three postulates: (a) vacancies are everywhere in thermal equilibrium; (b) cross diffusion terms (L_{ij} for $i \neq j$ in Eq. (1.5) are negligible); (c) plastic deformation in the diffusion zone is homogeneous. None of these assumptions can be considered self-evident and, in fact, postulates (a) and (c) do not appear to be confirmed by experiment. The restrictions become less important at lower concentration gradients. It seems clear that a diffusion couple consisting initially of pure metals involves too steep a concentration gradient for Darken's equations to apply. Despite these difficulties, Darken's equations have proved extremely successful in correlating the experimentally observed chemical diffusion coefficients with the self-diffusion coefficients of the two constituents for the few systems studied accurately to date.

4.6.2. Correlations with Solute Concentration

As mentioned earlier, there is some evidence that a correlation exists between the diffusion coefficient and the solidus temperature in concentrated alloys. It is frequently found that with solute additions which lower the melting point, the diffusivities of both solute and solvent atoms increase exponentially with increasing solute concentration, whereas solute additions which raise the melting point produce a corresponding decrease in diffusivities. In addition, those solutes which lower the melting point invariably diffuse more rapidly than do the solvent atoms themselves.

An exact atomistic treatment of the problem of diffusion in concentrated alloys has not yet been attempted. Before this can be done, a general statistical analysis is required to relate the diffusion coefficient to the many and varied jump probabilities

involved in an uncorrelated jump. This is a formidable problem even for dilute alloys, as shown by the recent work of Reiss (1959, R3). Some fairly drastic assumptions are necessary in order to make the problem manageable. Nonetheless, a number of useful relations have been formulated. Possibly the most significant is that proposed by Lidiard and LeClaire (1956, R3) who have related the diffusion of a tracer to the motion of the nearest-neighbor atoms. The results are generally consistent with a vacancy mechanism for diffusion. It appears that the presence of solute atoms changes the jump frequency of those solvent atoms which are mutual nearest neighbors of the solute and vacancy. This is an area in which careful and detailed calculations may be extremely fruitful.

4.7. Volumes of Activation

A proportionality between the experimental values of ΔV (Eq. (3.7)) and Q is found both for metals and ionic crystals, although the correlation is far from perfect. The correlation has been refined and improved by considering ΔV to be proportional to the product of the compressibility and activation energy (Keyes, 1958, R3). No direct information regarding the diffusion mechanism can be derived from this correlation, since calculations based on several different mechanisms yield relations of the same form. Furthermore, because of the approximations involved, the proportionality constant cannot be calculated with any degree of accuracy from first principles.

5. PROBLEM AREAS

5.1. Kirkendall Effect and Chemical Diffusion

The ideal conditions assumed in the Darken treatment of chemical diffusion, in terms of which experiments are generally interpreted, are experimentally approached only in diffusion couples in which very small concentration differences are present. Generally, marked departures from ideality exist, and these abnormal conditions are probably responsible for the differences between results still found in measurements of chemical diffusion coefficients on the same system by different investigators or methods. In absolute value the differences are not gross. In fact, the observed $c-x$ curves and marker shifts do not show any significant departure from the $t^{1/2}$ kinetics expected under ideal conditions, even at the shortest times observed; nevertheless, the discrepancies are at the same time appreciable, and show up particularly in the

derived activation energies. Too few systems have been examined in any detail and compared by different methods to reveal which of the various phenomena associated with the Kirkendall effect are most important in producing the variability of results observed, and systematic studies in this area are needed.

Considerable progress has been made in recent years in understanding the qualitative features of Kirkendall phenomena. However, the quantitative aspects are too poorly known to allow any good estimate to be made of the corrections which must be applied to the measured values of D . In any event, accurate values are more likely to come from improvements in experimental technique designed to avoid the complicating effects, and effort to this end is required.

The various phenomena associated with the Kirkendall effect have themselves attracted a good deal of study. Work has centered on the nature of the sources and sinks for vacancies, the mechanism of formation and growth of pores, the extent of the vacancy supersaturation, and the details of the plastic deformation processes, in particular the number of new short-circuiting paths produced.

It now seems clear that dislocations, by climbing, must be the primary sources and sinks for vacancies. This fact is evidenced, for example, by the comparative insensitivity of results to grain size or sample size. The way in which the operation of dislocations leads to the mass flow and lateral changes observed, and generates the stresses that lead to the plastic flow, seems to be fairly well understood, at least qualitatively.

Several methods have been employed to estimate the vacancy supersaturation, R , prevailing in the diffusion zone. For example, an estimate of the sink density and a measure of the net flux of vacancies into a given region, calculated from the observed concentration gradients and a knowledge of the intrinsic diffusion coefficients, serve to give rough values for R ranging from 0.01 to 100 at the positions where pores just form in α -brass, the highest value being for very pure material (Balluffi, Resnick, and Siegle, 1957, R3). Values for R of about .01 have also been estimated for Cu-Ni alloys from measurements of the critical pressure required to eliminate voids of known size already formed (Barnes and Mazey, 1958, R3). These and other estimates are always sufficiently large to indicate that dislocations could in fact operate continuously as sinks for vacancies, particularly under stresses as large as the yield stress. But equally important is the fact that the supersaturation is very much less than would be necessary to

nucleate pores homogeneously. For this reason, and because of the observed marked decrease in porosity in carefully purified materials, it is now believed that pores are nucleated heterogeneously, presumably at impurity particles or inclusions. Little is known about the details of the nucleation process or of the properties of particles most effective in catalyzing nucleation, although oxide particles are suspected responsible in α -brass.

None of the estimates of R are in any way precise, and more accurate values, as well as information concerning the way in which R varies with time and position, are much needed. The vacancy supersaturation necessary to nucleate pore formation is a known function of particle size, so that a direct measurement might possibly be made by studying the development of porosity in samples containing impurity particles of known amount and size (e.g., sols), but otherwise extremely pure.

An important quantity in discussions of R is the vacancy lifetime which is determined by the sink density. This lifetime is not at all well known, and the uncertainty here is reflected in some of the estimates of R . Quenching experiments suggest sink densities at low temperatures consistent with each point on a dislocation being an effective sink, giving lifetimes of 10^6 – 10^7 jumps. However, at the high temperatures of ordinary diffusion experiments, only direct encounters of vacancies with a jog might be effective in trapping and annihilating the vacancies, and those arriving at dislocations at points remote from jogs may readily reevaporate before they have time to reach a jog by diffusion along the dislocation. This effect might reduce the effective sink density by as much as a factor of a thousand. On the other hand, in chemical diffusion an increase in dislocation content, and therefore in sink density, is produced within the diffusion zone on account of plastic flow. If the vacancy density is sufficiently high, a further increase in sink density may arise from the generation of dislocation rings by vacancy aggregation and collapse. Clearly, a direct measurement of sink density or vacancy lifetime at high temperatures is needed. This measurement might be achieved by determining the time required to establish an equilibrium vacancy concentration in samples suddenly brought to a high temperature by pulse heating. Diffusion experiments in which extremely short annealing times are employed, though difficult, would be very informative, since the time to reach $t^{1/2}$ kinetics would be clearly related to the lifetime of a vacancy, and the degree of supersaturation would be manifest as an abnormal value of D . From the fact that $t^{1/2}$ kinetics are obeyed for the range of diffusion times so far

employed, Fara and Balluffi (1958) put an upper limit on vacancy lifetimes at 10^{11} jumps, for $R \sim 1\%$. Very short time experiments also provide a means for direct experimental distinction between vacancy and interstitial diffusion. If a vacancy mechanism were operative, abnormally high values of D would be found on the side losing atoms; for interstitial diffusion, high D values would be obtained on the opposite side of the interface.

The relative proportions of vacancies which end their lives at pores and at dislocations are determined in part by the ease with which dislocations climb. If climb is difficult, pore formation will be favored, and vice versa, other things such as nuclei density being equal. Little attention has been given to the effects of such differences in dislocation properties, for detailed study has been made of too few systems. Copper-based alloys have come in for far too large a share of attention, and the low stacking-fault energy of Cu is conducive to extensive dissociation, thus making climb difficult. In aluminum, climb should occur much more readily, and more detailed studies of systems based on Al or similar high stacking-fault energy metals might therefore be very informative. Other aspects of the Kirkendall phenomena might also depend on the deformation characteristics of the diffusing materials, particularly the propensity for lateral dimensional changes and the extent of crystal degradation; these features deserve further study.

5.2. Relation Between Chemical and Self-Diffusion

Development of means for determining chemical diffusion coefficients, in particular the intrinsic diffusivities, to an accuracy comparable with that of self-diffusion data is sorely needed in order to elucidate the relation between chemical and self-diffusion. Darken's thermodynamic relation appears to be moderately successful in accounting for current data. The fact that the model is theoretically invalid when there is any departure from equilibrium vacancy concentrations shows that such departures cannot be gross in practical cases. Precise data would provide a more critical test of Darken's relation, or of subsequent modifications which include the effects of non-equilibrium vacancy concentrations (Bardeen, Herring, 1951, R3) or of the Onsager cross terms appearing in the fuller thermodynamic derivation (LeClaire, 1959, R4).

All of these models attribute the differences between chemical and self-diffusion coefficients to differences in the thermodynamic "driving force" for diffusion, and pay little or no attention to the

detailed atomic processes. The essential problem is to understand the specific influence of a chemical concentration gradient on the normally random movements of atoms characteristic of self-diffusion. For self-diffusion, the net displacement of an atom, \bar{X} , is random in the sense that the average value of $\bar{X} = 0$. Due to the asymmetry introduced by the chemical gradient, this is no longer necessarily true in chemical diffusion, as shown by recent measurements of \bar{X} , or the "drift" (Manning, 1959, R3, Shuttleworth, 1959, R4). General relations between chemical and self-diffusion coefficients involving \bar{X} may be derived from random walk considerations which avoid the objections to a purely thermodynamic approach. The results are comparable with Darken's model in their agreement with experiment (LeClaire, 1959, R4). More extensive measurements of \bar{X} would be very welcome, for these constitute useful additional information on chemical diffusion behavior not provided by normal diffusion measurements.

Darken's equations follow from this more general random walk relation on making the usual thermodynamic approximation in evaluating \bar{X} . There remains, however, the need for an atomistic interpretation of \bar{X} which avoids such approximations, and which takes proper account of the influence that the varying configurations around any migrating atom exert on its jump rate. The first step would be to consider the simpler case of self-diffusion in concentrated alloys, a proper theory of which is still lacking, and then to study the effect of the concentration gradient. One product of such an investigation would be a clearer idea of the significance, if any, of the cross terms between different atomic aspects in the Onsager formulation of the diffusion equations.

5.3. Diffusion in Grain Boundaries and Dislocations

Evidence for preferential grain boundary diffusion has been frequently obtained, but no systematic studies have been made to determine how diffusion along grain boundaries influences chemical diffusion. Effects of this kind may be important, as shown by the dependence of marker shifts on grain size in Au-Pt couples. This matter deserves further attention. However, considerable study of boundary diffusion itself has been made in recent years, and the general features are now fairly well established. The possibility of quantitative study was first provided by Fisher's analysis (1951, R3). Refinements to this method have been made by Whipple (1952, R4) and by Vorisov, Kzolikow, and Lyubov (1956). Only the boundary condition of constant surface concentration has, however, been treated, and there is need for exami-

nation of other cases which might be more nearly approached experimentally.

Only comparatively recently has it been appreciated that there is an enhancement of D in low-angle grain boundaries, and even in polygon boundaries. Studies on boundaries of known dislocation structure and spacing lead to values for the diffusion coefficient along individual edge dislocation lines or "pipes," D_p . It has been found that D_p is of the same order of magnitude as the grain boundary diffusion coefficient, D_{gb} , in boundaries of maximum misfit; the larger part of the variation of D_{gb} with boundary angle appears to be attributable to variations in dislocation density, for D_p is either independent of angle (Ag), or at least very much less dependent on it (Ni) than D_{gb} .

However, the Fisher analysis, by which such experiments are interpreted, approximates the grain boundary as a uniform slab of constant width σ ; within this region the diffusion coefficient is D_{gb} , and hence the analysis yields only the quantity σD_{gb} , which can be related only crudely to D_p and to the dislocation density. Clearly, a more refined analysis, taking account of the individual dislocation structure, would give a more realistic value for D_p . Both low-angle boundaries and boundaries of greater misfit which are no longer completely describable as dislocation arrays, require further mathematical attention before the variation of D_{gb} with angle can be fully understood, since Fisher's analysis is appropriate only to the highest angle boundaries. Even for such boundaries, however, little is known about σ , which might well differ appreciably from the one or two lattice spacings generally assumed in diffusion problems, and may even vary appreciably with angle. It may not be a very good approximation to regard grain boundaries as discrete or sharply-defined regions. The possibility that a gradation in properties may exist from the center of the boundary into the bulk crystal might well be examined in relation to diffusion phenomena.

In the light of the comparatively recent demonstration that diffusion may occur quite rapidly along dislocations, Hart (1957, R3) has pointed out that because dislocations are very closely spaced compared with the diffusion length in typical experiments, the measurements may well contain a contribution from diffusion along dislocations. According to Hart, this contribution would not upset the c - x relation expected for pure diffusion—except of course at very large diffusion distances, where penetration by dislocation diffusion alone has been demonstrated (Williams and Slifkin, 1958). Hart estimates this contribution as $\sim 5\%$ in

Ag at 600°C. Such effects should show up as a slight decrease in Q with decreasing temperature. Decreases of the required order of magnitude (~ 1 K.cal/mole over a range of 400°C) have been claimed to have been observed and attributed to this cause (Tomizuka, 1958, R3), but the evidence is as yet far from unequivocal.

In measurements of the rate of impurity diffusion in metals or salts, the effect may be enhanced, inasmuch as the concentrations of diffusant are extremely low; consequently, if there is any appreciable tendency for segregation at dislocations, dislocation diffusion will contribute to a larger extent than in proportion to the fraction of sites on dislocations (Mortlock, 1956, R4). Low solubility may be conducive to ready segregation, and this may explain the very marked curvature in Mackliet's data (1958, R3) for Co and Fe diffusion in Cu, since with Ni, which is completely soluble in Cu, no curvature is observed.

More work is clearly needed on the role of dislocations in diffusion measurements, particularly with respect to impurity diffusion, and the possibility of enhanced effects due to segregation. Studies should be made using bicrystals of known misorientation, as well as low-angle boundaries of pure tilt, pure twist, and more complex character. Simultaneous experiments on "single" crystals of widely different structural perfection would be interesting, and would serve to eliminate other possible sources of apparent variations in Q , such as a slightly temperature dependent D_0 . Obviously extreme precision of measurement will be necessary.

Little is known about the factors which determine grain boundary or dislocation diffusion rates, or about any differences between self- and chemical grain boundary diffusion processes, although a more rapid variation of D_{gb} with angle has been suspected for chemical diffusion. Kirkendall-like phenomena and segregation of diffusant probably play some part. Arkharov (1959, R4) certainly finds that impurities segregating to boundaries can significantly accelerate or retard the grain boundary diffusion rate of one element into another. Present mathematical analyses take no account of segregation, the boundary concentration being determined only by diffusion parameters.

What is known about boundary diffusion has come mostly from studies on metals, but the interesting fact that there is an enhanced diffusion at grain boundaries only of the anions in alkali halides (Laurent and Benard, 1959, R4) should encourage detailed study of boundary diffusion in a wider range of materials.

5.4. The Importance of Self-Diffusion Experiments

Most of the methods for studying self-diffusion have been applied to metals, and to an increasing extent to ionic crystals, particularly the alkali and silver halides. The recent advent of the isotope exchange method for measuring self-diffusion coefficients is particularly welcome for ionic solids, since it provides means for studying anionic diffusion rates, not always measurable by conventional tracer techniques. However, application of the precision methods developed for metals has not been extended to other materials as widely as might be wished, particularly, for example, to semiconductors and to oxides and sulphides. Extensive and detailed knowledge of diffusion processes in such materials is rather lacking, but is surely important, if only because of obvious practical applications. In many studies on oxides and sulphides, it is probable that too little attention has been given to equilibration of the material in the atmosphere in which diffusion measurements are to be carried out (Birchenall, 1958, R1). To a lesser degree, this oversight may be true also of electrical conductivity measurements. Determinations of both diffusivity and conductivity should be carried out under conditions of constant partial pressure or of constant composition, or preferably both. Too often, powder compacts have been used, simply pressed together, where tracer deposition on single crystals would be far preferable, although it must be admitted that not all oxides or sulphides are readily available in single crystal form. Time devoted to the preparation of single crystals of these and other materials would be profitable, for it is perhaps only with well-characterized crystals that the substantial progress made in our understanding of diffusion processes in metals and ionic salts can be fully matched or extended to other materials.

5.5. Vacancy Configurations: Experiments Under Hydrostatic Pressure

A vacancy is only very incompletely described as a normal lattice site unoccupied by an atom. Thus, a very important question arises concerning the extent to which the absence of an atom influences the positions of surrounding atoms—important not only in considerations of the details of the diffusion process, but also in theoretical calculations of the energy and entropy of defects, of the interactions between them, of the X-ray and optical effects they produce, and of the resistivity associated with such defects. Several theoretical calculations of the relaxation of the lattice around a vacancy in close-packed metals have predicted values in

the range $\frac{1}{2}$ to 1 atomic volume for the volume change per vacancy formed. Recent calculations (Girifalco and Streetman, 1959) for body-centered cubic metals, where the nearest neighbors of a vacancy are not neighbors of one another, confirm the expectation that relaxation around a vacancy in this structure should be larger than that in face-centered cubic metals. Lomer (1958) has pointed out the very large relaxation which occurs around vacant sites in the equilibrium γ -brass structure. In the calculations on metals which have been made thus far, an equal relaxation on the part of each neighbor has been assumed. This restriction is not at all necessary, and less symmetrical relaxations might be considered, as indeed occur in the γ -brass structure. Calculations have also been made for ionic crystals of the NaCl type (e.g., Fumi and Tosi, 1959, R4). Here, an unsymmetrical type of relaxation almost certainly occurs, nearest neighbors moving away and second neighbors moving towards the vacancy.

All such calculations are far from precise and leave much to be desired. A principal difficulty is the lack of any satisfactory expression for the force law between atoms, and results are naturally sensitive to any assumptions made about this term. The difficulty is particularly acute because it has to be assumed that any chosen law of force and the constants it contains will apply under the abnormal conditions prevailing in the immediate neighborhood of the defect, although the constants have been evaluated by fitting the force law to data, such as elastic constants, which are characteristic of the perfect lattice. That this can indeed be a very inadequate assumption is well illustrated by the shortcomings of the commonly employed Born-Mayer expression for the exchange interaction between ion cores. Such an expression, fitted as it usually is to the elastic constants, appears to be quite inadequate to describe recent results on the pressure dependence of these constants, at least for Cu, Ag, and Au (Daniels and Smith, 1958), even for strains much less than those prevailing in situations involved in diffusion. In ionic salts, the high degree of polarization of the ions by the electric field of the defect they surround may further contribute to the inadequacy of a Born-Mayer expression. All defect properties calculated employing Born-Mayer potentials require re-examination in the light of these results, and there is a need for an alternative and more realistic expression for closed shell repulsions. Another serious defect is the common neglect of the electronic redistribution in the vicinity of the defect.

Experimentally, information on the extent of the relaxation may be derived from measurements of the pressure dependence

of diffusion coefficients. For substitutional diffusion, the observed pressure dependence indicates that ΔV (Eq. (3.7)) is positive, with a magnitude of about half an atomic volume. This is a somewhat smaller value than might be expected, bearing in mind that $\Delta V = \Delta V_i + \Delta V_m$. However, recent measurements for interstitial solute diffusion in iron yield a very small or zero value for ΔV ($= \Delta V_m$), while measurements for interstitial diffusion in vanadium gives values for $\Delta V \sim 0.2$ atomic volumes. If ΔV_m is small for vacancy diffusion, the observed values of ΔV are due largely to ΔV_i , and are then in agreement with Tewordt's (1958, R3) estimate of this quantity. Clearly, there is a need to determine ΔV_i and ΔV_m separately. Quenching experiments under pressure might be practicable, but any theoretical approach would require a much more detailed understanding of the processes of activated jumps than is currently available. For the present, the evidence is that ΔV_i is certainly less than the atomic volume, indicating appreciable inward relaxation of neighbors.

Experiments on the effects of pressure on diffusion are not the simplest to carry out, particularly when high temperatures are needed to achieve diffusion. The recent application of internal friction methods to the problem (Tichelaar and Lazarus, 1959, R3), making possible measurements at lower temperatures, provides a means for collecting data on a number of alloy systems.

Suggestions about the diffusion mechanism, involving considerations of the vacancy configuration, have come from experimental observations dealing with the correlation between diffusion and melting phenomena. Particularly relevant is Nachtrieb's (1954, R3) concept of "relaxions". Atomic rearrangement within a relaxion will contribute to any measured value of a chemical or self-diffusion coefficient. But the internal rearrangement will not contribute to any diffusion-controlled effects which require a net transport of defects, like marker shifts, sintering, microcreep, etc. Only the net migration of the relaxion will enter here. Calculation of such effects in terms of measured diffusion coefficients, using a simple vacancy model of one atom jump for one defect jump, would then give too large an estimate for the density of defects.

There is no evidence for such a two-process diffusion in systems for which relevant data exist. However, speculations of this kind serve to illustrate the fact that departure from the simplest concept of a vacancy as a point defect may have important consequences, and such possibilities are worth watching for, particularly in materials where appreciable relaxation might be expected. In this connection, it is interesting to note that recent experi-

ments on the hitherto comparatively neglected body-centered cubic metals and alloys show a diffusion behavior that departs in many respects from that characteristic of face-centered cubic systems, namely activation energies which are much lower than expected from their melting points (γ -U, Cr) together with rather low D_0 values (Adda and Kirianenko, 1959; Paxton, Pasierb and Kunitake, 1959).

5.6. Some Properties of Defects Measured in Equilibrium: Lattice Expansion

A particularly valuable approach is the method of determining c_v , the vacancy concentration, by comparison of dilatometric and X-ray measurements of the expansion coefficient (Simmons and Balluffi, 1959). Values of ΔH_v are frequently estimated by assuming theoretical values of ΔS_v , but could be obtained without this assumption, and ΔS_v itself determined, if techniques could be developed to give the vacancy concentration accurately as a function of temperature. Work of this type on body-centered cubic metals would be of special interest, for the apparently low activation energies for self-diffusion which are sometimes observed might be associated with low ΔH_v values, thus implying high and more easily measured vacancy concentrations.

Accurate density measurements, combined with precision lattice constant determinations, are also of value in determining the nature of those defects which must be present, in order to maintain electrical neutrality, in ionic crystals "doped" with multivalent impurities—a device of great value in studying diffusion processes in such materials. Interstitials contribute to an increase in density, whereas vacancies lead to a decrease, and the comparatively large defect concentration (roughly equal to the impurity concentration) makes experimental observation quite feasible; by comparison, extreme precision is required to obtain accurate measurements of the equilibrium defect concentrations in metals.

5.7. Properties of Defects Measured Under Non-equilibrium Conditions

Experiments on materials quenched from high temperatures so as to retain vacancy concentrations representative of the quench temperature have proven very valuable (Bauerle and Koehler, 1957, R3). The most commonly studied property is the excess resistance, $\Delta\rho$, due to the vacancies, but the excess energy released on annealing, ΔE , and the change in volume, ΔV , have also been determined. Measuring any of these properties as a function of

the quench temperature makes it possible to evaluate ΔH_r , and, when this quantity is known, ΔE then gives the defect concentration, c , and ΔS_r . Otherwise c and ΔS_r can be calculated only when theoretical values of $\Delta\rho/c$ or $\Delta V/c$ are employed. Clearly, there are obvious advantages in measuring all these quantities, $\Delta\rho$, ΔE and ΔV . This has been done so far only for gold; the values found for $\Delta\rho/c$ and $\Delta V/c$ provide important confirmation of other experimental or theoretical estimates. For example, knowledge of $\Delta\rho/c$ is particularly useful in estimating excess vacancy concentrations in radiation damaged materials. Quenching experiments in general provide a major source of information concerning the energies of formation of vacancies and, with estimates of ΔS_r , lead to values of 10^{-3} – 10^{-4} for c near the melting point for pure metals.

Very little experimental data are available on ΔS_r and ΔS_m separately. A theoretical calculation requires essentially a determination of the changes in frequency of the lattice vibrations due to the presence of the vacancies, or of atoms at saddle points. The reliability of these calculations is subject to the same limitations as appear in the computation of the relaxation, a knowledge of which is indeed necessary for the calculation. In addition, present calculations have employed only an approximate description of the perturbation of the vibrational spectrum. The more thorough approach which Montroll (1955, R3) has made to this problem is very welcome, since it reveals the possibility that defects may introduce localized vibrational modes with frequencies lying above the range of the continuum normal modes. This fact will surely be of considerable relevance in more exact calculations. For metals, the calculated entropy changes are of the same order as those found experimentally, but for sodium chloride there appears to be a considerable discrepancy in both sign and magnitude between theory and experiment, possibly due to the effects of polarization.

Since $Q = \Delta H_r + \Delta H_m$, the energy of motion, ΔH_m , can be evaluated when Q and ΔH_r are known. Alternatively, ΔH_m can be measured independently by studying the temperature dependence of the rate at which the quenched-in defect concentration decays on annealing. Such measurements usually form part of any quenching experiment, and the annealing is generally followed by resistance measurements, or, in substitutional solid solutions, by determining the increase in the relaxation time for stress-inducing ordering (Nowick, 1953, R3), as vacancies disappear.

Most experiments show that, consistent with theoretical estimates, ΔH_r is greater than ΔH_m in pure metals, and that the sum of the two is essentially equal to Q for self-diffusion. In the most

carefully conducted quenching experiments (Au), the agreement between the sum, $\Delta H_i + \Delta H_m$, and Q is excellent. In alloys, ΔH_i appears to be less than ΔH_m , possibly because ΔH_i decreases on alloying more rapidly than ΔH_m ; however, a satisfactory explanation of this interesting effect has not yet been advanced.

Valuable information regarding the nature of vacancy sinks may also be obtained from detailed studies of the annealing rate. The average number of jumps which an individual vacancy makes during the time required to anneal out (vacancy lifetime) is consistent with dislocations being the major sinks. When high concentrations of vacancies are quenched in from high temperatures, the initial annealing behavior is such as to suggest that the sink density were increasing during the annealing; this behavior has been interpreted as due to the formation and growth of additional dislocations loops by vacancy aggregation and collapse. (Kuhlmann-Wilsdorf, 1959, R4). Other measurements have also provided evidence for the trapping of vacancies at impurities, indicating that great care is needed in interpreting the results of quenching experiments.

Quenching experiments are clearly of considerable value in providing information on a number of important quantities. To date no quenching experiments have been carried out on body-centered cubic materials, although these might be most informative. However, though quenching experiments are simple in principle, results of useful precision are difficult to obtain, as the long history of the experiments on gold (1952-1957) testifies. These detailed studies have revealed the complexity, in particular, of the annealing process. It is thought that this complexity may be associated with the important part played by divacancies and possibly even larger sized clusters of vacancies, but several features remain to be explained. Much scope exists for further work, and additional measurements on ΔV and ΔE would be especially welcome.

Quenching experiments are possible also with ionic crystals, and experiments of this type have occasionally been performed. However, normal diffusion or conductivity measurements on crystals "doped" with impurities provide in this case a simpler means of resolving ΔH_i and ΔH_m .

Excess concentrations of vacancies may be introduced into a sample by techniques other than quenching, in particular by energetic particle irradiation and by cold work, and studies of the kinetics of their annealing behavior can give information on ΔH_m and on sink densities. The interpretation of such experiments is

not quite as straightforward as it might appear, however, because of the simultaneous generation of interstitial defects and dislocations.

Excess vacancy concentrations produced by any of these processes should be manifest as an increase in the diffusion rates, and experiments to detect such an increase have been carried out. With plastic deformation, however, part of any observed increase might come from rapid diffusion along the many extra dislocations and grain boundaries produced.

However, conventional diffusion experiments generally show little enhancement of D in materials after cold work or after irradiation; presumably, at the temperatures required, any excess vacancies or dislocations anneal out in a time period which is short compared with the time required for macroscopic diffusional flow to occur. But the rate of any micro-diffusion effect completed well within the lifetime of a vacancy may easily be affected by the excess vacancy concentration. For example, changes in degree of order (Cu_3Au , CuZn) have been observed after neutron or electron irradiation, at temperatures where thermally-induced changes occur too slowly to be measured experimentally.

Conventional diffusion experiments conducted during irradiation are similarly unsuccessful. At the high temperatures required, and with the presently available flux intensity, the steady excess vacancy concentration, Δc , is apparently not appreciably larger than the thermally-generated concentration, c . But as the temperature is lowered, c decreases while Δc increases, and the influence of Δc in enhancing diffusion rates can be seen as an increase in the rate of such micro-diffusion effects associated with changes in the degree of order.

The data obtained from measurements of D during plastic deformation are at present rather confusing. When an enhancement is observed, it is generally, as might be expected, proportional to the strain rate, but there is considerable disagreement concerning the magnitude of the effect. In Ag, results from different laboratories (Lee and Maddin, 1958; Darby, Tomizuka and Balluffi, 1959, R3) have been reported which differ by two or more orders of magnitude for similar strain rates and temperatures. The cause of the discrepancy is not known, and further work is obviously needed.

Experiments on diffusion during plastic deformation are of relevance to an understanding of the complex phenomena which occur within the diffusion zone during normal chemical diffusion, and for this reason alone more precise data would be welcome.

Apart from this, experiments on diffusion enhanced by radiation damage or deformation are probably of principal interest in studies of the irradiation or deformation processes themselves, for the effects observed depend primarily on the rate of generation of defects and their modes of decay.

Thermal diffusion measurements at one time seemed to offer promise as a means of determining ΔH_i and ΔH_m separately. However, the heat of transfer—the central quality in the phenomenological description of thermal diffusion—requires for an understanding of its sign and magnitude a very detailed consideration of the energy exchanges that occur during an activation jump. Specifically, the model must describe the way in which the activation energy is acquired, and the manner in which it is dissipated or transferred back to the surroundings. Current theories of the activated rate process are not sufficiently refined for this purpose.

For this reason, and in connection with many other diffusion problems (for example the isotope effect, activation volumes, activation entropies, etc.), there is considerable need for a more satisfactory theory of the jump process to replace the quasi-thermodynamic transition state expressions that are commonly employed. A useful model should pay particular attention to such features as the kinetic behavior of atoms surrounding the migrating atom, the influence of the defect on the local lattice vibrations, electronic effects, and quantum effects such as the zero-point energy. Recent papers by Vineyard (1957, R3) and Rice (1959, R3) represent welcome steps in this direction, but the problem is far from solved.

Thermal diffusion has acquired some technological importance in recent years because of the very steep temperature gradients that may arise in the components of nuclear reactor fuel elements. The extensive theoretical and experimental efforts now being brought to bear on this problem promise to provide valuable information on the nature of the activated process, the importance of which will extend well beyond any immediate practical applications. Recent experimental and theoretical work on thermoelectric phenomena suggests that studies of thermoelectric properties may form a useful adjunct to standard thermal diffusion experiments.

5.8. Interactions of Vacancies with One Another: Vacancy Pairs

It has been appreciated for some time that an associated pair of vacancies in a metal might have a much higher mobility than an

individual vacancy because of the reduced ion core repulsions experienced by an atom jumping into one of the pair. However, no convincing experimental evidence has been provided to show that divacancies do contribute measurably in normal diffusion experiments. Therefore, it must be concluded that either the binding energy is small and few divacancies are present in thermal equilibrium, or that their mobility is in fact much the same as that of a single vacancy. Nevertheless, in quenched material, the concentration of divacancies might be expected to increase relative to that present before quenching, and the changes in the annealing kinetics with the temperature of quench (in Au and Al) have been ascribed to just this effect. However, while estimates of the binding energy from quenching experiments (about 0.1 to 0.2 eV) are about the same as several theoretical estimates that have been attempted (about 0.3 eV), the corresponding estimate for ΔH_m is only 0.1 or 0.2 eV less than that for a single vacancy (~ 0.8 eV) which is surprising. If divacancies had a much lower ΔH_m , they would not be retained for any appreciable time, and so could not be invoked to explain the observed annealing behavior. A clearer understanding of the properties of divacancies would be most desirable.

Some low temperature recovery stages encountered during the annealing of irradiated or cold-worked metals might also be due to divacancies, but this is quite hypothetical and hardly confirms their existence. The activation energies associated with these recovery stages are, however, obligingly low (about 0.3 eV).

Pairs of vacancies of like kind are unlikely to form in ionic crystals because of the strong electrical repulsions between them. But vacancies of unlike kind attract as shown by Tosi and Fumi (1958), and effects which can be attributed to such pairs have been observed not only in diffusion and conductivity experiments, but also in connection with their optical and dielectric properties. Particularly relevant is the fact that, while vacancy pairs in ionic solids contribute to diffusion, they do not contribute to conductivity, because their net charge is zero. Lidiard (1957, R4) points out that studies of anionic diffusion, particularly below the "knee" in the extrinsic region, may offer the most fruitful source of information on divacancy diffusion.

5.9. Interactions of Vacancies with Impurities: Impurity Diffusion

Vacancies may interact with impurities because of the elastic or electrical forces between them. An electrical attraction between vacancies and solute impurities which are more electropositive

than the solvent has been invoked to explain the fact that such impurities diffuse with a lower activation energy than that corresponding to self-diffusion of the solvent (Lazarus, 1954, R3); a similar attraction explains the simultaneous enhancement of the solvent diffusion rates (Hoffman, Turnbull and Hart, 1955, R3). For less electro-positive solutes, the Q for impurity diffusion is higher than that for solvent self-diffusion, consistent with a repulsion between vacancies and impurities. The small observed differences in Q suggest a small binding energy and indicate only a temporary association of vacancy and impurity. This contrasts with the more or less permanent vacancy-solute complexes which were at one time postulated to explain early, but now discredited, experimental data.

These results have stimulated some theoretical work on the electronic interaction between vacancies and impurities in metals (Fujiwara, 1958, Friedel, 1954, March, 1957, Blatt, 1955, R3). The importance of the subject warrants even greater attention. Work has until recently been based on the Thomas-Fermi approximation, and the resulting interaction energies appear to account quite well for the experimental data on solute diffusion, at least in systems for which z , the excess valency of the impurity, is positive. However, more recent and exact studies of the electron distribution around impurities suggest oscillations in the sign of the potential and of the excess electron density which are not apparent in the simpler treatments. Since the first nodes appear at distances of the order of inter-atomic spacings, it is clear that some considerable modification of the elementary theory of solute diffusion may be necessary. All theoretical work so far has been confined to impurities in noble metals. Systems in which z is negative present difficulties to the Thomas-Fermi model and a satisfactory treatment of such systems is notably lacking. The behavior of impurities in transition metals, where the neglect of departures from the assumed free electron behavior becomes serious, would also merit consideration in view of the increasing amount of data now becoming available on such systems.

Comparatively little attention has been given to quantitative studies of the elastic interaction between vacancies and impurities, entailing consideration of size effects, ionic deformability, etc., and more effort in this direction would be welcome. Both elastic and electronic interactions will no doubt need to be considered in any fully satisfactory theory of impurity diffusion.

Evidence for the association of vacancies and impurities in ionic crystals is provided by Chemla's demonstration that divalent

cation impurities in NaCl diffuse much more rapidly than sodium ions and with a lower activation energy. This vacancy-impurity association is confirmed by the fact that an electric field produces a displacement of the impurities which is far less than that expected from their diffusion coefficients, showing them to be present largely as neutral entities. Other evidence for the association of vacancies with impurities or with host ions comes not only from similar comparisons of diffusion and conductivity data, but also from studies of dielectric relaxation (Haven, Dryden, and Meakins, 1958, R4) and of electron spin resonance phenomena (Watkins, 1959, R4). The latter experiments (Mn in NaCl and KCl) even distinguish isolated Mn atoms, Mn atoms with an associated vacancy in a nearest-neighbor position, and with a vacancy in a next-nearest-neighbor position.

5.10. Correlation Effects: Mass Dependence of Jump Rates

A strong association between vacancies and impurities will enhance the correlation effect in vacancy diffusion first pointed out by Bardeen and Herring (1951, R3). The value of the correlation factor f in Eq. (3.1), and therefore the value for D solute, depends on the rates of exchange of the vacancy not only with the impurity but also with neighboring and next-nearest neighbor solvent atoms (LeClaire and Lidiard, 1956, R3). This effect complicates the interpretation of observed solute diffusion rates, which measure only a mean or effective jump frequency, because of the difficulty in determining the larger number of relevant variables.

The problem is to determine either the correlation factor f itself, or, ideally, the separate jump frequencies which determine its value. Information regarding pertinent jump frequencies might be forthcoming from studies of the diffusional narrowing of the nuclear magnetic resonance lines of the solvent and solute atoms as a function of temperature, since the line width is determined by the rates of atomic motion in the immediate neighborhood of the resonating nuclei; alternatively, the associated spin-spin relaxation times could be measured. A measure of the correlation factor itself can be obtained by comparing D solute for different isotopes of the solute species.

Further experiments on isotope effects in impurity diffusion, and indeed generally, will be of considerable interest, but call for extreme precision if results of useful significance are to be obtained. After all, we are looking here for accurately measured differences in D of the order of a few percent, whereas mere

repeatability of results within this accuracy is usually a matter for self-congratulation

The present interpretation of isotope experiments is based on the result, derived from the current treatments of activated processes, that the jump rate of an atom of mass m is proportional to $m^{-1/2}$. The interest in isotope experiments has encouraged a reexamination of this approximation. Vineyard's (1957, R3) more careful treatment has already suggested that jump rates might well vary more slowly than in proportion to $m^{-1/2}$. There is a great need for further experimental investigation in this area. The recent measurement of the rates of diffusion of interstitial C^{12} and C^{13} in Fe by a magnetic method (Bosman, Brommer, and Rathenau, 1959) show the $m^{-1/2}$ law to be obeyed to within $\pm 25\%$, confirming earlier but less accurate measurements on a few other systems. However, the high accuracy necessary to establish the law with any degree of firmness has hardly yet been achieved. In metals, experiments on interstitial diffusion or self-diffusion in high-purity systems are the most direct; in ionic salts advantage may be taken of Chemla's (1959, R4) technique of magnifying the isotope effect by carrying out the diffusion experiment in an electric field. Confirmation or modification of the $m^{-1/2}$ law will be of value in checking the more sophisticated theories of jump processes.

If and when accurate experimental values of the correlation factor become available, they will provide means of investigating atomic jump processes in metals and alloys at a new level of detail. This is already occurring in ionic crystals, where f can be determined directly by a comparison of diffusion coefficients and ionic conductivities. An excellent example is the demonstration by McCombie and Lidiard (1956, R4) that Ag diffuses in AgCl by an interstitialcy rather than by an interstitial mechanism. Many similar possibilities for exploiting the importance of correlation effects in diffusion are likely to occur in future work. The correlation factors have been calculated for a number of lattice structures and diffusion mechanisms (Compaan and Haven, 1958), but experimental work of the required precision has been largely confined to materials having the NaCl structure. Only by making accurate measurements on a wider range of materials can full advantage be taken of those calculations.

One correlation problem that has not yet been satisfactorily solved is that involving the effect of solute-vacancy pairs on the diffusion coefficient of the solvent. In view of the availability of

reliable experimental data, some effort towards a calculation of f would be welcome.

It is unfortunate that ionic conductivity measurements cannot be made as readily on alloys as they can on ionic salts, so that f can be determined in a similar fashion. Alloys can be electrolysed, but the ionic mobility does not seem to be simply related to the diffusion jump probability, apparently because of the strong influence exerted by the electronic current. Such effects are nevertheless of great interest and warrant further study.

5.11. Conclusions

To summarize, it is clear that no aspect of the subject of diffusion can be regarded as in any way exhausted. The general appeal is for more accurate measurements, and for their extension to a wider range of materials. On the former depends the extent to which the finer and more fascinating details of atomic diffusion processes can be elucidated, and the problems awaiting solution in isotope effects, chemical diffusion and grain boundary and dislocation diffusion are examples illustrating the need. Extension to a wider range of materials is essential if a comprehensive knowledge of solid state diffusion is to be acquired, for the features understood at present, derived from study of a comparatively restricted range of solid types, are surely not representative of the whole.

Among metals and alloys, only those with close-packed structures have been studied in any detail; body-centered cubic systems are beginning to attract attention and show interesting behavior, but the more complex metal structures, intermetallic phases, etc., are comparatively unstudied. Among ionic salts most attention has been given to NaCl-type structures, with almost complete neglect of others. In nonmetallic materials showing some electronic conductivity—semiconductors, oxides, sulphides, etc.—the intriguing detail of diffusion and defect behavior revealed by the very few systems that have been studied fully should encourage extension of such work to other systems. Of the diffusion behavior characteristic of the welter of other inorganic and organic materials, practically nothing systematic is known, outside of gas permeation phenomena.

Diffusion experiments are comparatively simple in principle, and so are within the reach of most normally equipped laboratories, but the measurements do require extreme care and precaution

at all stages if useful results are to be obtained, from the very choice and preparation of material onwards. "Excellent agreement" between the results of experiments on the same system by different investigators is still a rarity, and any effort devoted to rectifying this situation will be of inestimable value to the further progress of the subject. Recent developments in means of purification and of crystal growth, prompted by their need in semiconductor applications, can now be used to good advantage in all diffusion work. Particularly important, too, is that proper attention should be given to controlling all relevant variables in diffusion experiments. A sad example here is the not uncommon neglect of providing carefully controlled atmospheres in many experiments on oxides etc., leading to uncertainties in the exact composition, and therefore in defect concentrations. Valuable and oftentimes indispensable information can also be obtained by measurements of other physical properties such as conductivity, lattice parameter, dielectric loss, and so on, as well as the dependence of the diffusion coefficient on atmosphere, composition and pressure in relevant cases. It is obviously important that wherever possible the same material be used for these indirect or supplementary measurements as for the diffusion experiments themselves. The practice is common in studies on salts but should be more widespread.

Theoretically too, as we have seen, much remains to be done, but it is easier to point to shortcomings than to say what should be done about them. Theoretical calculations on diffusion phenomena are difficult, rarely free of considerable approximation, and, therefore, limited in their application—a feature common to most branches of solid state theory. The important part which theories play is in establishing the models, the framework, the language in terms of which the observed phenomena are described and discussed, and in this way profitably influence the course of experimental investigation. Many "theories" only have their day, then gently fade away; but good experimental facts always remain, and it is only on the basis of these that any subject can be firmly built. Sufficient is now known to design and conduct meaningful experiments of good accuracy, and in the present state of the subject, the most important immediate need would seem to be to exploit this fact to the fullest, not only in conventional diffusion experiments, but also in the many indirect means of investigation that are finding increasing application.

6. REFERENCES

Review Articles

- R1. Birchenall, C. E., *Metallurgical Reviews*, 1958, *3*, 225.
R2. Hauffe, K., 1955, *Reaktionen in und an Festen Stoffen* (Springer, Berlin).
R3. Lazarus, D., 1960, in *Advances in Solid State Physics* (Academic Press, New York), Vol. 10, p. 71.
R4. LeClaire, A. D., 1949, in *Progress in Metal Physics* (Pergamon Press, London and New York), Vol. 1, p. 306; 1953, Vol. 4, p. 265; Vol. 10 (to be published).
R5. Seith, W., and Heumann, T., 1955, *Diffusion in Metallen* (Springer, Berlin).
R6. Tomizuka, C. T., 1959, in *Methods in Experimental Physics* (Academic Press, New York), Vol. 5.

Recent Articles not Included in Review Papers

- Adda, Y., and Kirianenko, A., *Jour. Nuclear Materials*, 1959, *2*, 120.
Bosman, A. J., Brommer, P. E., and Rathenau, G. W., *J. Phys. Rad.*, 1959, *20*, 241.
Compaan, K., and Haven, Y., *Trans. Faraday Soc.*, 1958, *54*, 1498.
Daniels, W. B., and Smith, C. S., *Phys. Rev.*, 1958, *111*, 713.
Fara, H., and Balluffi, R. W., *J. Appl. Phys.*, 1958, *29*, 1133.
Girifalco, L. A., and Streetman, J. R., *Phys. Chem. Sol.* (in the press).
Lomer, W. M., 1958, in *Vacancies and Other Point Defects in Metals and Alloys* (Inst. of Metals, London).
Paxton, H. W., and Kunitake, T., *Trans. Metallurgical Soc., AIME* (to be published).
Paxton, H. W., and Pasierb, E. J., 1960, *Trans. Metallurgical Soc., AIME* (in the press).
Schoen, A. H., *Phys. Rev. Letters*, 1958, *1*, 138, 184.
Simmons, R. O., and Balluffi, R. W., 1960, *Phys. Rev.* (in the press); 1959, *Phys. Rev. Letters*, *3*, 577, 578.
Sippel, R. F., *Phys. Rev.*, 1959, *115*, 1441.
Tosi, M., and Fumi, F. G., 1958, *Nuovo Cimento*, *7*, 95.
Turnbull, D., and Hoffman, R. E. (to be published).
VorISOV, V. T., Kzolikow, V. M., and Lyubov, B. Ya., *Izvest. Akad. Nauk. SSSR, Otdel. Tekh. Nauk.*, 1956, *10*, 37.
Williams, G. P., and Slifkin, L. M., *Phys. Rev. Letters*, 1958, *1*, 243.

Part VI

PHASE TRANSFORMATIONS IN THE SOLID STATE

Chairman

Dr. MORRIS COHEN

*Massachusetts Institute of Technology
Cambridge, Mass.*

Members

Dr. JOHN W. CAHN
*General Electric Co.
Schenectady, N. Y.*

Dr. J. W. CHRISTIAN
*University Museum
Oxford, England*

Dr. PAUL A. FLINN
*Westinghouse Electric Corp.
Pittsburgh, Pa.*

Dr. MATS HILLERT
*Royal Institute of Technology
Stockholm, Sweden*

Dr. L. KAUFMAN
*Manufacturing Laboratories,
Inc.
Cambridge, Mass.*

Prof. THOMAS A. READ
*University of Illinois
Urbana, Ill.*

CONTENTS

	Page
1. INTRODUCTION	312
2. THE CHEMICAL DRIVING FORCE	315
2.1. Statement of the Problem	315
2.2. State of Knowledge	316
2.3. Current Theoretical Ideas	317
2.3.1. Random Configurational Entropy of Mixing	317
2.3.2. Lattice Electronic Energy (0°K) of Mixing	317
2.3.3. Vibrational Free Energy of Mixing	319
2.3.4. Magnetic Free Energy of Mixing	320
2.3.5. Electronic Specific Heat Contributions	320
2.3.6. The Lattice Stability of Metals	321
2.4. Suggestions for Future Work	322
3. PRE-TRANSFORMATION PHENOMENA	323
3.1. Statement of the Problem	323
3.2. State of Knowledge	323
3.3. Current Theoretical Ideas	324
3.4. Suggestions for Future Work	325

	Page
4. CRYSTALLOGRAPHY	326
4.1. Statement of the Problem	326
4.2. Current Ideas Concerning Lattice Geometry and Types of Interfaces	326
4.2.1. Relations between Lattices	326
4.2.2. Coherent Interfaces	327
4.2.3. Semicoherent Interfaces	327
4.2.4. Crystallography of Martensite	330
4.2.5. Diffuse Interfaces	331
4.2.6. Steps in Boundaries: Nonplanar Interfaces	332
4.3. Current Ideas Concerning Surface and Strain Energy	332
4.3.1. Surface Energy	332
4.3.2. Strain Energy	333
4.4. Crystallographic Features of Transformations	334
4.4.1. Widmanstätten Structures	334
4.4.2. Martensitic Transformations	334
4.4.3. Bainitic and Other Intermediate Transformations	335
4.4.4. Order-Disorder Transformations	335
4.4.5. Discontinuous Precipitation and Eutectoid Reactions	335
4.4.6. Noncrystallographic Transformations	335
4.4.7. Crystallography of Growth	335
4.5. Critical Questions and Suggestions for Future Work	335
5. ORDER-DISORDER AND CLUSTERING REACTIONS	336
5.1. Order-Disorder Reactions	336
5.1.1. Statement of the Problem	336
5.1.2. State of Knowledge	337
5.1.3. Current Theoretical Ideas	338
5.1.4. Critical Issues and Suggestions for Future Work	339
5.2. Clustering Reactions	340
5.2.1. Statement of the Problem	340
5.2.2. State of Knowledge	340
5.2.3. Current Theoretical Ideas	341
5.2.4. Critical Issues and Suggestions for Future Work	342
6. NUCLEATION	343
6.1. Statement of the Problem	343
6.2. State of Knowledge	343
6.3. Current Theoretical Ideas	344
6.3.1. Limits of Metastability	344
6.3.2. Evaluation of the Work of Forming a Critical Nucleus	346
6.3.3. Evaluation of the Nucleation Rate	347
6.4. Other Aspects of Nucleation	348
6.4.1. Two-Stage Nucleation	348
6.4.2. Auto-catalyzed Nucleation	348
6.4.3. Two-Dimensional Nucleation as a Growth Mechanism	348
6.5. Suggestions for Future Work	348
7. THE GROWTH PROCESS	349
7.1. Statement of the Problem	349

PHASE TRANSFORMATIONS IN THE SOLID STATE 311

	Page
7.2. Current Knowledge and Ideas Concerning Boundary Movements	350
7.2.1. Coherent Glissile Interfaces	350
7.2.2. Glissile Semicoherent Interfaces: Fast and Slow Growth	351
7.2.3. Single Interface Transformations	353
7.2.4. Cessation of Growth: Stabilization	354
7.2.5. Thermally Activated Motion: Factors Affecting Mobility	355
7.2.6. Theory of Thermally Activated Growth	357
7.2.7. Critical Questions Concerning Boundary Movements	358
7.3. Current Knowledge and Ideas Concerning Long-Range Atom Transport	358
7.3.1. Diffusive Transport Through Bulk Phases	359
7.3.2. Diffusion Short Circuits	359
7.3.3. Mathematical Solution of the Diffusion Equation	359
7.4. Other Factors in the Growth Process	360
7.4.1. Role of Impurity or Alloying Elements	361
7.4.2. Role of Coherency Stresses	361
7.4.3. Role of Mechanical Stress and Pressure	361
7.4.4. Role of Interfacial Energy	362
7.5. Current Knowledge and Ideas Concerning the Kinetics of Growth	362
7.5.1. The Rate-Controlling Step	363
7.5.2. Lamellar Growth in Precipitation and Eutectoid Decomposition	364
7.5.3. Bainitic Growth	365
7.6. Suggestions for Future Work	367
7.6.1. On Boundary Movements	367
7.6.2. On Atomic Transport	368
7.6.3. On Complicating Factors in Growth Process	368
7.6.4. On Kinetics of Growth	368
7.6.5. On Lamellar Growth	369
7.6.6. On Bainitic Growth	369
8. GENERAL KINETICS OF SOLID-STATE TRANSFORMATIONS	370
8.1. Statement of the Problem	370
8.2. State of Knowledge	371
8.2.1. Use of Classical Rate Laws	371
8.2.2. Evaluation of Rate of Nucleation and Rate of Growth	371
8.2.3. Phenomenon of Impingement	372
8.2.4. Significance of Activation Energy	373
8.3. Current Ideas	375
8.4. Suggestions for Future Work	376
9. SUMMARY	376
10. BIBLIOGRAPHY	380
10.1. Review Articles	380
10.2. References	380

PHASE TRANSFORMATIONS IN THE SOLID STATE

1. INTRODUCTION

The science of phase transformations should enable one to predict the time-temperature conditions under which a reaction will occur in a given system. In addition, it should be possible to foretell the nature of the transformation and to calculate its overall kinetics. In the final analysis, it should even be possible to describe in advance the ultimate morphology resulting from the transformation, and thereby control the structure-sensitive properties of materials.

It goes without saying that the current science of phase transformations falls far short of achieving this goal. There is now strong need for carefully controlled experiments and realistic theories if we are to make substantial progress toward a deeper understanding of phase transformations. The main purpose of this report is to set the stage for such progress.

The relevant subject matter has been arranged in systematic order, starting with the driving force for transformations and the state of the parent phase prior to transformation. This is followed by a discussion of the crystallography of phase transformations, including the nature of interfaces and the evaluation of strain energy. Ordering and clustering reactions are then considered, followed by treatments of nucleation and growth processes, and of general kinetics. In each of these subjects, there is an attempt to highlight (a) the essential problem, (b) the state of knowledge, (c) the main theoretical ideas, and (d) important directions for future work. An overall summary is given at the end.

Certain limitations of this study should be emphasized at the outset. Although morphology is undoubtedly an integral aspect of phase transformations, an adequate approach to this subject has not been found. This situation results from the fact that morphology is dependent upon many contributing phenomena which are not sufficiently understood in themselves. Prior history, fine-scale variations in the structural state of the parent phase, preferred nucleation sites, competitive growth processes, impingement, short-circuiting diffusion paths, all combine to affect morphology in a major way. Consequently, the more basic phenomena are given attention here, in the hope that morphology can be treated more adequately at some future date.

It is probably a reflection on the provincialism of the authors that the present report is based primarily on metals. Although

the principles of phase transformations are undoubtedly applicable to all materials, the fact remains that some nonmetallic materials may lend themselves especially well to important experiments, improved techniques, or simplification of overlapping factors. Transformation studies on semiconductors and transparent halides, for example, may contribute to the understanding of phase transformations in a profound way. However, the authors found it desirable to concentrate on metals: (1) because it is on metals that the largest body of systematic knowledge is available, and (2) because even in this well-studied class of materials, the critical problems require elucidation. This approach should help to identify the key experiments that might be done with nonmetals, and it is likely that there will be due attention to this aspect of the problem in the future.

The usual schemes of classifying solid-state transformations have become unsatisfactory because of the increasing number of specific reactions that "do not fit." Accordingly, a new system of classification is suggested here, as outlined in Table 1. An initial distinction is made depending upon whether the transformation is homogeneous or heterogeneous. A homogeneous transformation is considered to be one which takes place everywhere within the parent phase without the formation of new interfaces. Clustering and ordering reactions of the second kind are well-known examples. However, most of the transformations of interest fall in the heterogeneous category, i.e., new interfaces are created. It should be noted that even if a transformation is initiated by homogeneous random nucleation, the transformation is classified as heterogeneous in the terminology adopted here because new interfaces are involved.

The subclassifications under Heterogeneous Transformations in Table 1 focus attention on the nature and kinetics of the growth process. We distinguish here between shearlike processes which give rise to shape changes (or surface tilts) and those which are not accompanied by any shape change although a volume change may be operative. Growth in the latter group is characterized by the movement of an incoherent interface, while growth in the former group is characterized by the movement of a highly coupled interface which may be either fully-coherent or semi-coherent. In a fully coherent interface, there is complete matching of crystallographic planes and rows of atoms between the parent phase and the transformation product across the interface. Differences in atomic spacing are accommodated by elastic strains.

In a semicoherent interface, part of the misfit in atomic spacing is accommodated by dislocations.

The growth processes are further categorized by whether or not the interface motion is thermally activated. In nonthermally activated growth processes, the interface is glissile and responds sensitively to the net driving force. Since thermally activated atomic jumps or relaxation phenomena are not required, this type of interface movement is inherently fast. On the other hand, when thermal activation controls the growth process, time as well as temperature become important factors in determining the interface motion. Nevertheless, even for thermally-activated growth, the interface may be fully or semicoherent and result in a shape change, or it may be incoherent and produce no shape change. Examples of both are indicated in Table 1.

It may be noted that the suggested classification does not depend upon the nature of the nucleation process, whether homogeneous or heterogeneous, isothermal or athermal, random or nonrandom. This is a reflection of our lack of knowledge concerning the fine-scale details of nucleation phenomena. Instead, it is more useful to adopt a classification having an operational basis, and therefore the growth process seems to provide the best distinguishing features of transformations at the present time.

In view of the related study on "Growth, Structure and Morphology of Crystals," liquid-solid transformations and crystallization from the amorphous state are not covered here. Solid-state recrystallization, magnetic changes and ferroelectric transitions are also regarded beyond the scope of the present inquiry.

It is a pleasure to acknowledge the valuable aid received from the following men who read the preliminary version of this report and offered important suggestions: B. A. Bilby, R. Bullough, M. E. Fine, R. F. Hehemann, L. Himmel, T. B. Massalski, G. R. Speich, R. A. Swalin, D. Turnbull, and C. A. Wert. The detailed comments of L. Himmel were particularly helpful. Special thanks are also due to J. J. Harwood of the Office of Naval Research for his vision in initiating and supporting the program.

TABLE 1

Classification of Solid-State Transformations

- I. Homogeneous transformations (no new interfaces formed)
 - A. Clustering (increase in number of like atomic neighbors)
 - B. Ordering of the second kind (increase in number of unlike atomic neighbors)
 - 1. Short-range ordering
 - 2. Long-range ordering

II. Heterogeneous transformations (accompanied by formation of new interfaces)

A. With attendant shape change (shear-like macroscopic displacements, surface tilts)

1. Nonthermally activated growth—glissile interfaces (regular martensitic transformations)
2. Thermally activated growth—nonglissile interfaces
 - a. Controlled by short-range atomic transport¹ (some polymorphic transformations like Widmanstätten ferrite; the disordered fcc → ordered orthorhombic reaction in Au-Cu; isothermal growth in AuCd under certain conditions)
 - b. Controlled by long-range atomic transport² (bainite; probably precipitation in Cu-Be)
 - c. Controlled by relaxation processes (probably low-temperature isothermal growth in U-Cr)

B. Without shape change (presumably this type of growth is always thermally activated).

1. Controlled by short-range atomic transport¹ (polymorphic transformations which are not shear-like, in contrast to IIA2a; order-disorder reactions of the first kind in contrast to IB; massive transformations as in Cu-Ga)
2. Controlled by long-range transport² (precipitation; lamellar growth, including discontinuous precipitation and eutectoid decomposition).

2. THE CHEMICAL DRIVING FORCE

2.1. Statement of the Problem

In order to predict the conditions under which a phase transformation may occur in a given system, a knowledge of the chemical driving force for the reaction is required. Customarily, this driving force is described in terms of the difference in bulk free energy between that of the transformation product and that of the parent phase. However, there has been some question concerning the applicability of macroscopic thermodynamics to the nucleation of small particles with dimensions of a few atomic spacings. Unfortunately if bulk thermodynamic data are not valid for the fine-scale phenomena involved in phase transformations, or if the reaction paths comprise atomic configurations that differ from the parent and product arrangements, then a detailed analysis must be postponed until the thermodynamics of all structures can be generally derived from first principles. The possibilities of performing such calculations are quite remote at present.

In spite of these reservations, the results obtained thus far in

¹ Short-range atomic transport signifies movements by only a few jump distances; for example, across an interface or atomic interchange within an interface. Composition changes do not occur.

² Long-range atomic transport signifies movements of many jump distances. Composition changes do occur.

predicting the temperature and composition ranges of transformations have been rather promising (Turnbull, 1956), (Kaufman and Cohen, 1958), (Kaufman, 1959a). Hence, it seems probable that the uncertainties introduced in using macroscopic thermodynamic data are no worse than the corresponding errors in kinetic measurements or in strain and surface-energy computations.

The evaluation of a driving force requires an explicit description of the free energy/temperature/composition surfaces for the parent and product phases, each of which may be stable, metastable or unstable depending upon temperature and composition. Direct measurements of the driving force can sometimes be made in the regions of composition/temperature space where the phases are stable or metastable. Evaluation of the free energy of unstable phases requires extrapolation techniques which usually depend upon splitting the free energy contributions into physically meaningful terms. Although there is at present no generally applicable theory of solid-solution thermodynamics, several semi-empirical treatments have been suggested which specify the various contributions to the free energy of mixing by means of a limited number of parameters (Oriani, 1957, 1958R; Weiss and Tauer, 1958a). These approaches attempt to utilize physical measurements to calculate the thermodynamic properties. Although the methods are too simple in most instances, they do provide some connection between the physical and chemical aspects of solution thermodynamics. This connection appears to be essential for an explicit description of free-energy surfaces which, in turn, are necessary for driving-force computations.

2.2. State of Knowledge

From a quantitative standpoint, determination of the various thermodynamic functions of solid solutions by calorimetric and physicochemical techniques is far more advanced than the corresponding theory. This situation prevails despite the inherent difficulties and low precision in some of the measurements. Experience indicates that the extreme values of the enthalpy of mixing are about ± 3000 cal/mol for solid solutions, while the excess entropy (over that of an ideal solution) usually lies between $\pm 1.5R$ cal/mol $^{\circ}K$.

Thermodynamic or statistical-mechanical theories cannot predict the magnitude, or even the sign in many cases, of the excess entropy and enthalpy of mixing. However, various contributions to these functions have been recognized and formulated in terms

of measurable quantities. Although the evaluation of any one of the contributions does not solve the overall problem, it does bring us one step closer to an explicit description of the free energy/temperature/composition surface.

2.3. Current Theoretical Ideas

Contributions to the free energy of mixing which are due to atomic configuration, lattice electronic energy, vibrational entropy, strain energy, magnetic free energy, and changes of the electronic specific heat on alloying have been recognized. In most instances, it is not possible to predict with certainty which of these contributions will be important in any given system. For some cases, however, it is possible to estimate the relative importance of some of these contributions if certain physical properties of the system in question have been measured.

2.3.1. Random Configurational Entropy of Mixing

The configurational entropy of a random substitutional solid solution has been calculated, and the results substantiated to some extent on specific metallic systems (Lumsden, 1952). The random or ideal entropy has become the "norm" to which solid solutions are compared. This is the only contribution to the entropy in ideal solutions.

The ideal entropy has also been computed for interstitial solid solutions for which geometrical considerations indicate a given number of interstitial sites per host-lattice atom (Johannson, 1937). Unfortunately, accurate thermodynamic data for interstitial solutions are rare and the corresponding deviations from ideality have not been adequately investigated.

2.3.2. Lattice Electronic Energy (0°K) of Mixing

A variety of statistical, cellular, and electronic models have been developed in order to predict the sign and magnitude of the electronic energy, or to relate this term to one or more physical parameters. These models have had some limited success when compared qualitatively with experiment (Oriani, 1957, 1958R).

In the zeroth approximation of the quasi-chemical model, the energy is set proportional to $x(1 - x)$, where x is the atomic fraction of one component in a binary phase and the solution is considered to be random. The statistical models which are based on central-force interactions attempt to calculate the enthalpy and excess entropy of solution, as well as the deviation from atomic randomness of the solution, in terms of an Ising parameter. For

an ordering alloy, the statistical model relates the local configuration, the anomalous specific heat, and the free energy of mixing, to the ordering temperature. The primary thermodynamic contribution of the ordering is to the energy, with the excess entropy providing only a small negative contribution (usually less than 0.3 cal/mol °K). However, comparison with experiment does not yield uniformly encouraging agreement (Oriani, 1958R). It is not always clear whether the difficulty lies with the assumption of nearest-neighbor interactions or with the inadequacy of the mathematical approximations introduced in solving the statistical problem. (It has been suggested that anisotropic terms, second and higher-neighbor interactions, as well as terms leading to composition-dependent interaction parameters, are significant in metallic solutions and cannot be neglected (Oriani, 1957, 1958R)). Since an analytic solution to the three-dimensional Ising problem is not available, the statistical theories are based upon approximations which are not reliable when there are strong interactions. Numerical techniques designed to give better approximations to the problem have yielded consistent results for the 50/50 β -brass structure, and lend some support to the utility of a nearest-neighbor interaction parameter (Flinn, McManus, and O'Meara, 1959).

Application of these techniques to solubility-gap problems could conceivably provide an important insight into the details of local atomic clustering prior to phase separation, and might also shed considerable light on the nature of the nucleation problem.

It should be recognized that the interaction parameter, and the associated energy and entropy described by the statistical models, are short-range in nature and arise only from nearest-neighbor correlations. On the other hand, the experimentally determined enthalpy measures the sum of long-range and short-range terms. For example, if the experimentally determined enthalpy of mixing in a given system is approximately $Ax(1-x)$, where A is a positive constant, the constant may be composed of a small negative short-range term and a predominantly large positive term leading to a miscibility gap despite the presence of short-range order. It is possible that AuNi provides an example of this type of behavior (Flinn, Averbach, and Cohen, 1953). However, the existence of short-range order in this system, has been disputed by recent measurements (Munster and Sagel, 1958).

The 0°K enthalpy of mixing need not be parabolic or symmetrical with respect to composition. This situation has been treated semi-empirically with a subregular model (Hardy, 1953) in which

$$H = x(1 - x) (A_0 + A_1x) \quad (1)$$

where A_0 and A_1 are constants. Some success has been achieved in predicting the sign of the second derivative of H from electron theory (Friedel, 1954), while from an empirical point of view Eq. (1) has been used to describe thermodynamic data obtained experimentally (White, Orr, and Hultgren, 1957).

In addition to the energy contributions discussed above, the misfit energy due to the disparity in size between the constituent atoms in solution seems to be an important factor in the formation of metallic solutions. Although considerable effort has been devoted to various elastic sphere-in-hole models, it has been shown recently (Oriani, 1958R) that the effect of out-of-size atoms is manifest in liquid as well as solid solutions. Hence, the elastic calculations cannot be the correct procedure for deducing the associated energy and entropy attributable to the misfit atoms. No reliable method of computing this misfit energy and entropy is available at present.

2.3.3. Vibrational Free Energy of Mixing

Evaluation of the vibrational free energy, arising from the changes that occur in the vibrational spectrum on alloying, requires an accurate determination of the vibrational spectrum of each atomic species as a function of composition and temperature. The primary contribution to the excess free energy of an alloy will come from the entropy. If the vibrational frequency of the alloy is approximated by an average Debye temperature, $\theta(x)$, then in a binary system A-B containing an α solid solution, the excess vibrational entropy, $S_E^\alpha(\frac{\theta}{T})$, is defined as

$$S_E^\alpha\left(\frac{\theta}{T}\right) = S\left(\frac{\theta(x)}{T}\right) - (1-x) S\left(\frac{\theta_A}{T}\right) - x S\left(\frac{\theta_B}{T}\right) \quad (2)$$

In Eq. (2), θ_A^α and θ_B^α are the average Debye temperatures of pure A^α and pure B^α . The sign and magnitude of the excess vibrational entropy depends upon the $\theta^\alpha(x)$ curve. Limiting values are probably of the order of $\pm R$.

Since the $\theta^\alpha(x)$ curve is important in determining $S_E^\alpha\left(\frac{\theta}{T}\right)$, it is desirable to couple calorimetric and physicochemical measurements with Debye-temperature measurements. However, coordination of such measurements has not yet been made. The $\theta^\alpha(x)$ curves can be determined by X-rays (James, 1954R), elastic constants (DeLauney, 1956R), or specific heat measurements (Blackman, 1955R).

Any fundamental electronic theory capable of predicting the elastic constants of alloys together with the energy terms considered in section 2.3.2 would probably be capable of predicting $\theta^\alpha(x)$. At present, however, no such theory exists, and the $\theta^\alpha(x)$ curve must be determined experimentally. In the absence of other means of evaluating $\theta^\alpha(x)$, it might be useful to estimate $\theta^\alpha(x)$ on the basis of the Lindemann formula (Blackman, 1955R).

2.3.4. Magnetic Free Energy of Mixing

The excess magnetic free energy of the phase α is defined as

$$F_E^\alpha(u) = F^\alpha(u) - (1-x) F_A^\alpha(u) - x F_B^\alpha(u) \quad (3)$$

where $F^\alpha(u)$, $F_A^\alpha(u)$, and $F_B^\alpha(u)$ are the magnetic free energies of the alloy and the pure components, respectively, at a given temperature. Approximate equations have been developed (Weiss and Tauer, 1958b), relating the magnetic free energy to the lattice structure and the Curie or Néel temperatures. For a ferromagnetic f.c.c., solution at temperatures well above its Curie point

$$F(u) \approx -RT \left(\frac{T_c(x)}{1350} \right) \quad (4)$$

If $T_c(x)$ is known, the excess magnetic free energy of mixing can be estimated as a function of temperature and composition. Comparison of Eq. (3) and the approximate $F(u)$ relations (Weiss and Tauer, 1958b) with activity measurements in systems exhibiting magnetic transitions, or in systems where the magnetic contributions may be dominant, would be useful in testing the approximate formulations.

Magnetic contributions are quite prominent in some transition-metal systems. The excess magnetic entropy may be as high as $\pm R$ while the 0°K enthalpy may be ± 1000 cal/mol.

2.3.5. Electronic Specific Heat Contributions

If the electronic specific heat coefficient $\gamma^\alpha(x)$ is known for the α phase of a binary system, then the corresponding excess entropy of mixing is approximately

$$S_E^\alpha(\gamma) \approx [\gamma^\alpha(x) - (1-x) \gamma_A^\alpha - x \gamma_B^\alpha] T \quad (5)$$

The corresponding enthalpy is

$$H_E^\alpha(\gamma) \approx [\gamma^\alpha(x) - (1-x) \gamma_A^\alpha - x \gamma_B^\alpha] \frac{T^2}{2} \quad (6)$$

The electronic specific heat coefficient can be accurately determined by low temperature specific-heat measurements (1° to 4°K) in metals and alloys where magnetic transitions are absent or can

be accounted for (Keesom and Pearlman, 1955R). Experimental determinations of electronic specific heat coefficients for pure metals indicate values of 0.1×10^{-4} to 5×10^{-4} cal/mol $^{\circ}\text{K}^2$ for nontransition metals and 1.0×10^{-4} to 40×10^{-4} cal/mol $^{\circ}\text{K}^2$ for transition metals. At present, corresponding data for alloy systems are rare. However, the results available for binary systems indicate that systematic measurements of $\gamma(x)$ curves may be valuable in guiding the development of the electron theory of metals in addition to providing important thermodynamic data (Rayne, 1957; Wei, Cheng, and Beck, 1959).

Equations (5) and (6) indicate the importance of the electronic specific heat terms at high temperatures. Under certain circumstances, $F_E^{\alpha}(\gamma)$, which is proportional to T^2 , may become a dominant contribution to the free energy of mixing. Inspection of (5) and (6) shows that if $[\gamma^{\alpha}(x) - (1-x) \gamma_A^{\alpha} - x \gamma_B^{\alpha}]$ is approximately $\pm 10 \times 10^{-4}$ cal/mol $^{\circ}\text{K}^2$, which is quite possible for transition-metal binaries, then at 1000 $^{\circ}$ K, $S_E^{\alpha}(\gamma)$ and $H_E^{\alpha}(\gamma)$ become $\pm 0.5R$ and ± 500 cal/mol respectively. It is evident that such contributions to the free energy are quite significant.

2.3.6. The Lattice Stability of Metals

Since it is necessary to define the free energy/composition/temperature surfaces for two (or more) phases in considering the chemical driving force, it is important to know the difference in free energy between any two possible allotropic modifications of the pure elements themselves. Schemes for describing this difference in terms of zero-point electronic and magnetic energies as well as of vibrational and electronic specific-heat contributions have been suggested (Weiss and Tauer, 1956). Encouraging progress in this area has been made on metals exhibiting phase transformations such as iron (Weiss and Tauer, 1956), manganese (Weiss and Tauer, 1958b), and titanium and zirconium (Kaufman, 1959a). In addition, it may be possible to obtain similar information for metals which do not exhibit allotropy by analyzing systems for which the phase diagram and the free energy of mixing of the phases involved are known (Lumsden, 1952; Kaufman, 1959b).

The advantage of having lattice-stability data for the pure elements is that this information can be applied to any system involving these elements. Therefore, such information has a wider range of direct applicability than the data for the free energy of mixing in a single system. Presently, there is no way of predicting the stability of atomic configurations in metals with

sufficient precision to permit reliable estimates of the free-energy difference between various structures.

2.4. Suggestions for Future Work

(a) Progress toward a quantitative understanding of the free energy/composition/temperature surfaces, which is necessary for driving-force calculations, will require a coordination of theory and experiment as well as a coupling of physical and chemical measurements in which phase equilibria, calorimetry, activity data, diffraction and electronic studies are brought together. Section 2.3 indicates the relationships between the physicochemical and the physical properties of an alloy system. Thus, if the thermodynamic properties of the α phase of the A-B system were determined by chemical activity and calorimetric studies, it would be desirable to decompose the observed thermodynamic properties into the component parts. X-ray or neutron diffraction studies on the same system could be employed to determine the local atomic configurations and vibrational frequency of the alloys. Low-temperature specific-heat measurements could be used to determine the electronic specific-heat coefficient as a function of composition. If the system exhibits magnetic properties, measurement of the susceptibility as a function of temperature and composition would be desirable for evaluating the magnetic component of the free energy.

(b) In addition to information regarding the individual phases, α and β , of the A-B system, it is necessary to know relative positions of these phases in free energy/temperature/composition space. Advances in this direction will be facilitated by a quantitative knowledge of the lattice stability of metals and by an explicit representation of the free-energy surfaces of the individual phases. The application of calorimetric techniques to diffusionless transformations would be quite helpful in this connection. Experimental determination of the enthalpy change attending f.c.c. \rightleftharpoons b.c.c. reactions in Fe-Ni alloys (Normann and Scheil, 1958) has been very valuable in ascertaining the relative free energies of the f.c.c. and b.c.c. phases in this system as a function of temperature and composition. This technique could be profitably applied to other systems exhibiting diffusionless transformations involving metastable states (i.e. Co, Ti, and Zr-base phases).

(c) The current availability of high-pressure facilities may afford the opportunity of examining the properties of stable and unstable phases over a range of pressure, temperature, and composition. In addition to yielding important data on the com-

pressibility of stable phases, it may be possible to expand the composition and temperature fields in which direct observations can be made on a given phase. Furthermore, observations may become possible on phases that are normally unstable at atmospheric pressure.

(d) Experimental determination of the thermodynamic activities of interstitial components in solid solutions would help establish the validity of the ideal-entropy calculation for such phases, as well as elucidate the nature of other contributions to the thermodynamic behavior. This field of investigation is basic to an understanding of the extraordinary lattice stability of interstitial compounds, but seems to have been neglected. There are very few interstitial systems for which precise activity data are available. Titanium and zirconium-base systems, which exhibit large interstitial-phase fields based on the b.c.c. and h.c.p. lattices, appear ideally suited for such measurements.

3. PRE-TRANSFORMATION PHENOMENA

3.1. Statement of the Problem

A detailed knowledge of the structure of the parent phase is necessary for the full understanding of the mechanism and kinetics of a phase transformation. The equilibrium structure is related to the thermodynamic functions, and in some cases (order-disorder reactions, for example) important changes in fine-scale structure occur within the parent single-phase field which are inseparable from the transformation itself. Imperfections in the parent phase, such as vacancies, dislocations, and grain boundaries, may also play important roles in the transformation, both in nucleation and in the diffusion processes associated with growth. It is evident, then, that the exact nature of the parent phase, in all its fine-scale aspects, must receive thorough attention.

3.2. State of Knowledge

It now seems well established that solid solutions (in the general sense of variable-composition phases) are not usually random arrangements of the component atoms located at lattice points, but rather, are correlated arrangements of atoms slightly displaced from lattice points (Warren and Averbach, 1953). The correlation of the positions of substitutional solute atoms, and some measure of the displacements from lattice points (atomic-size effect) can be determined by X-ray diffraction techniques, and the necessary measurements have been carried out for several alloys. This entire field has recently been reviewed quite thor-

oughly (Guinier, 1959R). Interstitial-interstitial and interstitial-substitutional interactions have been measured by an internal friction technique (Powers and Doyle, 1959). The formation, motion, and condensation of vacancies have recently been measured in several materials, and some data are now available on the activation energies for formation and motion of vacancies (Kimura, Maddin, and Kuhlmann-Wilsdorf, 1959), and on the nature of the condensation products. However, this work has been mainly confined to f.c.c. metals. Dislocations, and their interactions with impurities, have been studied principally with regard to the mechanical properties of metals and alloys (Cottrell, 1954R) but some of the results are useful in analyzing phase transformations. Qualitatively, at least, it appears that strong interactions exist between dislocations and solute atoms, both substitutional and interstitial.

3.3. Current Theoretical Ideas

In many alloy systems, the correlation of positions of substitutional atoms seems to be isotropic, so that it can be associated with pairwise interactions between atoms, formally equivalent to bonds. If the ratio of the interaction energy to kT exceeds some critical value, the correlations cease to be only local and a phase transformation occurs; clustering, if the interaction between like atoms is attractive; ordering, if the interaction between unlike atoms is attractive. Formally, above a critical temperature

$$\lim_{n \rightarrow \infty} |\alpha_n| = 0 \quad (7)$$

and below the critical temperature

$$\lim_{n \rightarrow \infty} |\alpha_n| > 0 \quad (8)$$

The quantity α_n is defined as follows. Consider the difference between the number of unlike n th-neighbor pairs actually present in a solid solution and the number which would be present in a random solution of the same composition. The ratio of this difference to the corresponding number for a random solution is defined as α_n . If there is an excess of unlike neighbors, α_n is negative; if there is a deficiency, α_n is positive. The significance of this term is discussed further in section 5.

Various attempts have been made to associate the distortion of the lattice, due to differences in effective atomic size, with a "strain energy" but no thoroughly satisfactory results have been obtained, as pointed out in section 2.3.2. Even the analysis of the

diffraction data requires some sort of model, and not enough is known about interatomic forces to do this satisfactorily (Averbach, 1956R).

The interactions of interstitials with substitutional solute atoms or other interstitials are generally ascribed to lattice strain effects, and again a good quantitative treatment does not appear possible at present. The importance of vacancy condensates (loops, sheets of stacking faults, and helical dislocations) in the nucleation of precipitates has been shown (Nicholson, Thomas, and Nutting, 1959). It seems possible that, except in special cases, precipitates are normally nucleated at such heterogeneities. There is also evidence that strong attraction may exist between vacancies and substitutional solute atoms. This affects both the number of vacancies present (by lowering the effective energy of formation) and their mobility at low temperatures.

3.4. Suggestions for Future Work

(a) Direct confirmation of the adequacy of the simple direct interaction picture for substitutional solid solutions would be valuable. This could perhaps be obtained by a comparison of the interaction energy inferred from diffraction measurements of local order, with a calorimetric measurement based on the difference between fast (no atomic rearrangement) and slow (equilibrium order maintained) specific-heat measurements.

(b) Direct measurements of the association of interstitials with other interstitials and with dislocations may be possible by neutron diffraction, and might lead to very fruitful results. In particular, a comparison between the structure (determined by diffraction) and the energy interaction (determined by internal friction measurements) could be made. As in the case of substitutional interactions, the statistical-mechanical problem has not been solved exactly.

(c) Some improvement of our understanding of the energy of lattice distortion on an atomic scale is badly needed.

(d) Studies of vacancy behavior (as in quenching experiments) should be extended to other systems. In particular, no data yet exist for b.c.c. metals or for alloy systems.

(e) The effect of vacancies and dislocations on the degree of local clustering or ordering requires elucidation.

(f) Extensive studies by thin-film electron microscopy on the role of heterogeneities in nucleation should be pursued.

4. CRYSTALLOGRAPHY

4.1. Statement of the Problem

A significant aspect of solid-state transformations is the crystalline nature of the phases involved. This becomes especially evident whenever the structures have well-defined mutual orientations, or whenever the reaction product exhibits shape anisotropy. The crystallographic features often enter into the classification of phase transformations. However, from the standpoint of the present study, the essential point about the crystallography is that it sheds important light on the elementary atomic movements which occur during transformation. Inferences can also be drawn concerning the nature of the interface and the state of strain attending the reaction. These facets of crystallography are emphasized in the following sections. General reviews of this subject are given by Barrett (1952R), Bowles and Barrett (1953R), Bilby and Christian (1955R), and Christian (1956R).

4.2. Current Ideas Concerning Lattice Geometry and Types of Interfaces

4.2.1. Relations between Lattices

The crystallographic relation between two lattices may be expressed as one of an infinite number of affine transformations¹ in which one set of lattice points is converted or transformed into the other. These descriptions are all equally valid, unless the structures are coherent or semicoherent in such a way that one lattice is effectively transformed into the other as the boundary migrates. When this happens, a "correspondence" between the lattices may be defined, and one particular description has physical rather than only formal significance. Thus, the relations between two twins may be described as rotations or as a simple shear, but the latter is more significant when discussing a mechanical twin.

An affine transformation may be analyzed into a pure deformation combined with a pure rotation. A necessary condition for two lattices to fit together on a common plane is that they shall be related by a pure deformation in which one principal strain is zero, and the other two have opposite signs. Inasmuch as this implies a definite relation between lattice distances, the condition

¹ An affine transformation is one in which straight lines are converted into straight lines, and planes into planes. Most formal crystallography is expressed in terms of lattices, since the positions of atoms cannot be connected by an affine transformation in the general case. The structural relations are completed by "shuffling" the atoms within some unit cell (not necessarily a primitive cell).

can only be satisfied coincidentally, the only known practical examples being f.c.c. \rightleftharpoons h.c.p. transitions with nearly identical nearest-neighbor distances within the close-packed planes. It follows that the affine transformations relating two lattices cannot be such that there is an invariant plane of either rational or irrational type. The condition for a line to be invariant is merely that all three principal strains do not have the same sign, and this can almost always be satisfied.

4.2.2. Coherent Interfaces

Two fully coherent structures have rows and planes of atoms continuous across the interface, which contains no dislocations. Except for regions of a few atomic distances, two crystals can be fully coherent only across a planar interface, and only then if the lattices are such that the above condition for the existence of an invariant plane is very nearly satisfied. The boundary between the f.c.c. and h.c.p. planes in cobalt and in certain cobalt alloys is always plane and appears to be coherent, any slight difference in nearest-neighbor distances within the close-packed planes being accommodated elastically. Many twin boundaries are fully coherent.

Forced elastic coherence is possible if sufficiently small regions of one structure form in a defect-free region of another. The permissible strains are then very much larger than those at the normal elastic limit.

4.2.3. Semicoherent Interfaces

A boundary in which regions of continuity are separated by regions of misfit (e.g. dislocations) may be described as semicoherent. The difficulties of matching two structures over even one plane imply that semicoherent crystals will invariably be plate-shaped when embedded in a matrix. It is useful to distinguish two extremes, for which the name "epitaxial and martensitic semicoherent interface" are appropriate.

Epitaxial matching implies that the two structures involved have almost identical configurations or lattice points in planes parallel to the common (habit) plane. The densities of lattice points in these parallel lines are, however, different in the two structures, and the slight differences in spacings and angles have to be accommodated by interface dislocations with Burgers vectors in the interface plane. The simplest model is one in which two mutually perpendicular directions are parallel in the two structures, the matching being accomplished by a crossed grid of

edge dislocations. As such a boundary moves normal to itself, lattice sites are created on one side of the interface, or removed from the other, and there must thus be a mass flow to or from the boundary region. The habit plane will be rational in both lattices; it need not contain a direction which is invariant in the lattice change.

Epitaxial semicoherence might be expected, for example, when one cubic phase precipitates from another of nearly identical lattice spacing, but there is no definite experimental evidence for the existence of boundaries of this type. Precipitate phases often remain coherent with the matrix until the particles grow quite large, but the nature of the interface is unknown in detail, except for copper-beryllium alloys, where the semicoherency seems to be of the martensitic type.

Martensitic boundaries are plane in single-interface transformations and in some polycrystalline transformations. The more usual lenticular shape may be attributed to the stresses developed when the transformation is constrained by the surrounding matrix, and a clear analogy exists with the corresponding situation in mechanical twinning. There is a macroscopic shape deformation associated with the formation of a martensitic plate, manifested by the tilting of a polished surface about the line of intersection of the habit plane. In some martensitic structures, the product consists of a stack of twins which are visible under an optical microscope, and which produce surface tilts within the martensitic plate. Recent thin film electron microscopy (Kelly and Nutting, 1959) shows that martensitic plates in steel and in some other alloys are stacks of very fine twins, the minimum thickness observed being about 20 atom diameters. No evidence of a dislocation structure in the interface has yet been found.

Existing crystallographic theories (Wechsler, Lieberman and Read, 1953; Bowles and Mackenzie, 1954a, 1954b; Bullough and Bilby, 1956) assume that the habit plane is either exactly or approximately an invariant plane of the shape deformation. The shape deformation is obtained by combining a lattice and a lattice-invariant deformation, the latter being a simple shear. It follows that either the product plate consists of a stack of fine parallel twins, or that the dislocations in the interface are equivalent to a set of parallel lines with a common Burgers vector.

The glide plane of the dislocations, or the K_1 plane of the twins, is the "plane of second shear" of the crystallographic theories. "Corresponding" planes of this type in the two lattices are not parallel but meet edge to edge in the habit plane. The dislocations

are parallel to the line of intersection, which is (exactly or approximately) the invariant line of the lattice deformation. The Burgers vector of the dislocations need not lie in the habit plane, and in the published theories, it is assumed to correspond to a slip or twinning vector in both lattices. Rows of lattice points parallel to these corresponding vectors meet at an angle in the interface. The Burgers vector must then be described in a reference lattice, which may be either of the two real lattices or some median lattice; it is usually convenient to choose the parent lattice. The whole array of dislocations may be treated as a two-dimensional continuous distribution, and is then described as a glissile surface dislocation (Bullough and Bilby, 1956).

The discussion requires some modification when the martensitic region consists of a stack of fine twins. In the theory of Bowles and Mackenzie (1954a, 1954b, 1957), it is assumed that the twins have equivalent correspondencies with the parent, and it follows that the twinning plane is derived from a mirror plane of the parent. This allows alternate lamellae of the two twin orientations to end at the interface, and either of these may be regarded as the fundamental product orientation. The stack of twins is macroscopically equivalent to a single crystal martensitic plate of either orientation, although the lattice invariant shear, and hence the dislocation spacing, is different for the two possible single crystal products. If twins are formed, the shear specifies only the relative amounts of the two orientations, and the actual width of each twin must be determined by some additional condition, presumably an energetic condition.

The coherency strain energy at the interface can be minimized by having narrow twins, with widths of the order of the dislocation separation in the corresponding single orientation model. This may represent the situation in steels, and is generally to be expected if the transformation strain is large. With smaller strains, the energy of the coherent twin boundaries also becomes important and the equilibrium width is greater. The energy can probably be minimized by having narrow twins at the interface broadening into much wider twins behind the interface. This seems to happen in single interface indium-thallium transformations. In gold-cadmium alloys, however, the wide twins extend to the interface. Hence, there must be dislocations either in the interface, or possibly ahead of the interface, as indicated by the tilts sometimes produced in the parent phase.

4.2.4. Crystallography of Martensite

The predictions of the theory of martensite crystallography when the elements of the lattice invariant shear (slip or twinning plane and direction) are assumed to be rational are in good agreement with experiment in some transformations, but not in others. Extensions to the theory have been made in two ways. The first possibility is to retain the assumption that the habit plane is an invariant plane of the shape deformation, but to allow either the slip direction or the slip plane, or both, to be irrational. The second possibility is to allow small departures from the invariant plane strain condition.

Crocker and Bilby (1959) have investigated the effect of varying the slip plane and direction for the transformations in steel, uranium and titanium. The method of calculation developed by Bullough and Bilby (1956) was programmed for use with a digital computer, and they were thus able to cover a very large number of possibilities. The results show that experimental data can generally be explained in this way, but only if both the slip plane and direction are allowed to be irrational; there are a few remaining discrepancies; e.g. in the orientation relations for titanium and zirconium. The use of an irrational slip plane and direction is permissible if, for example, the interface contains more than one kind of dislocation. The requirement that the whole interface shall be a glissile surface dislocation seems to be met if either all the dislocations are parallel but have differing Burgers vectors lying in the same slip planes, or, alternatively, if the dislocation lines are nonparallel but have a common Burgers vector. The relative proportions of the two kinds of dislocations may differ from plate to plate, depending on the constraints of the matrix, and may thus explain the real scatter in habit planes which is observed experimentally (Otte and Read, 1957).

An alternative approach, suggested by Lieberman (1958), is to reverse the usual calculation, using experimental results as input data and thus determining the nature of the lattice invariant deformation. Unfortunately, the scatter and errors in the results seem to be too great for this to be feasible in most cases.

Bowles and Mackenzie (1954a, 1954b, 1954c, 1957) severely limit the possibilities for the lattice invariant shear by postulating that twinned orientations with equivalent correspondences can form in one plate, as is known experimentally to be true in some transformations. They obtain good agreement with experiment by supposing that there are small differences in the lengths of

large vectors parallel to the interface in the two lattices. This difference is uniform in their treatment; in Frank's model of {225} martensite in steels (Frank, 1953), a change in length in one direction only of the interface is assumed.

If the interface is not quite invariant over macroscopic distances, the crystallography is modified and the habit plane no longer contains the invariant line of the lattice deformation. This type of interface may be regarded as a combination of the semicoherent martensitic and semicoherent epitaxial types. Physically, it seems essential that the dilatation be sufficiently small to represent an elastic change of lattice spacing at the interface, as was also assumed by Frank. The only alternative is to introduce additional dislocations into the interface to compensate for the change in length, but it is very probable that these would destroy the glissile nature of the interface (section 7). As a small plate grows edgewise, increasing the interfacial area, the elastic stresses due to forced macroscopic coherence become large, and these might eventually force nucleation of edge dislocations in the interface. This might be the situation in stabilized martensitic plates in which the interface has become immobile.

At present, it does not seem possible to deduce the nature of the lattice invariant deformation from the experimental results in all cases, nor to decide whether or not the shape deformation always has an invariant plane. These problems are more likely to be solved by direct observation of the structure of martensite using transmission electron microscopy than by measurement of crystallographic parameters.

4.2.5. Diffuse Interfaces

Recent work by Cahn and Hilliard (1958) emphasizes the importance of diffuse interfaces under some conditions. The interfaces they consider are not crystallographic, but crystallographic structures do sometimes give a macroscopically diffuse interface. An example of the latter is the tapering of a set of parallel twin boundaries to eliminate one orientation and produce a single crystal slightly rotated from the other twin orientation. This situation is described as a distributed shear and is effectively a diffuse semicoherent boundary. An exactly similar boundary can occur in a " λ -type" interface transformation (Basinski and Christian, 1954a, 1954b). Apart from such special cases, however, most semicoherent boundaries are probably sharp rather than diffuse.

4.2.6. Steps in Boundaries: Nonplanar Interfaces

A step in a coherent twin boundary is described as a twinning dislocation. If the step height is h , and the twinning shear is s , the Burgers vector of the dislocation is $b=hs$, and is thus parallel to the direction of the twinning shear. Bilby (1953) has called the corresponding step in a coherent martensitic boundary a transformation dislocation. Its properties are similar to those of a twinning dislocation, but it has a Burgers vector with a component normal to the habit plane.

A curved interface cannot be completely coherent, but if the radius of curvature is large, it may be composed of relatively long coherent sections separated by atomic steps. A simple model of a lenticular plate comprises closed loops of transformation dislocations on successive planes parallel to the habit plane. The diameters of the loops decrease with increasing distance from the central plane of the plate.

A nonplanar semicoherent boundary may be treated in the same way. Consider a martensitic plate in the form of a disc of uniform small thickness. The plane of the disc is the habit plane, and the upper and lower surfaces each contain a set of parallel dislocations or twin boundaries. The edges of the disc will not be coherent, and if these are only a few atoms in extent, it is convenient to regard the dislocation loops as continuous across these regions as well. Thus, there are closed dislocation loops lying in the surface of the plate, the plane of the loops being normal to the habit plane. These loops are clearly a different nature from the loops of transformation dislocations considered earlier, which are parallel to the habit plane, and exist only at steps in the interface.

The example of a flat disc is used to emphasize the fact that the loops of dislocations normal to the habit plane have no essential connection with any curvature of the habit plane interface. One possible description of a lenticular martensitic plate is that of closed dislocation loops (or twin boundaries) normal to the habit plane, crossed by closed steps (transformation dislocations) parallel to the habit plane. If the steps are widely spaced, the loops of dislocations, or the twin boundaries, lie completely in the habit plane, except for atomic steps.

4.3. Current Ideas Concerning Surface and Strain Energy

4.3.1. Surface Energy

The structural part of the surface energy of coherent, semi-

coherent and incoherent interfaces may be considered in the same way as the energy of low and high-angle grain boundaries. For semicoherent interfaces, the energy is mainly that of the dislocation array, and Brooks (1952R) has indicated that formulae quite analogous to the Read-Shockley grain boundary equation should be valid. The surface energy should be small and orientation-sensitive for semicoherent interfaces, and large and orientation-insensitive for incoherent interfaces.

4.3.2. Strain Energy

When a particle forms inside a rigid matrix, an extra energy term is introduced because of the change in volume and shape of the transformed region. The whole assembly is in a self-stressed state, and the strain energy is proportional to the volume of the particle if the shape remains constant. The energy is, however, very dependent on shape and lattice orientation, and this may be especially important in deciding the easiest nucleation path.

Eshelby (1957) has described a set of virtual operations by which the strain energy for a coherent or semicoherent particle may be calculated, using the approximation of isotropic elasticity. Unfortunately, it does not seem possible to obtain a closed solution in a real lattice, except for the case of hexagonal symmetry. It should be noted that for almost all kinds of coherent and incoherent boundaries, the strain energy is minimized if the particle is plate-shaped. This tendency is opposed at the nucleation stage by the surface-energy factor.

In epitaxial coherence, the strain energy resides mainly in the particle, rather than in the matrix. A special case of some interest is that which exists when the stress-free (unconstrained) transformation is a pure dilatation. This is illustrated by the precipitation of one f.c.c. structure from another of different lattice parameter; several examples are known in which the lattices are parallel and the precipitate is a plate with a {100} habit. In the approximation of isotropic elasticity, the strain energy for coherent precipitates of this kind should be independent of the shape of the particle. This is an exception to the general rule mentioned above, and the formation of plates with a specific habit requires an explanation. The observed habit may be due either to (1) elastic anisotropy favoring {100}, (2) surface energy favoring {100}, or (3) kinetic growth factors.

For martensitic coherence per se, the strain energy is due only to the macroscopic shape deformation. In nucleation problems, it is not certain that the critical stage includes production of a mar-

tensitic type of semicoherent interface, but the available evidence seems to support this. The shear component of the invariant plane strain usually produces a large strain energy, most of which resides in the matrix.

The crystallography of martensite is largely fixed by the necessity of having a glissile interface which can remain semicoherent throughout the growing stage. Nevertheless it is possible that the introduction of some dilatation is required to reduce either the strain or the surface energy by varying the crystallography (Christian, 1958).

A somewhat different model is used to calculate the strain energy of an incoherent particle. Since there is no continuity of structure, atoms may be transferred from one region of the particle to another to minimize the energy. This is equivalent to filling a hole in the matrix with a compressible fluid of different natural volume. The strain energy of an incoherent particle is least for a plate, tending to zero as the plate thickness decreases, is a maximum for a sphere, and is reduced somewhat for a needle.

4.4. Crystallographic Features of Transformations

4.4.1. Widmanstätten Structures

These are the oldest known examples of crystallographic transformations. The lattice relations presumably indicate that the planes are coherent at the nucleation stage; either type of coherency is possible, but epitaxial coherence which usually involves the lower energy is more likely. The habit plane may be determined by an energy-minimization condition, or it may be determined by growth kinetics. Tilts of the plates are not generally expected; where observed they probably indicate martensitic coherence which persists until the plates are of considerable size. If tilts are found, the transformation is perhaps correctly described as bainitic (sec. 4.4.3).

4.4.2. Martensitic Transformations

These are the most fully crystallographic phenomena. Semicoherence is maintained throughout the growth. The factors determining the orientation relations at the habit plane have been summarized in section 4.2.4. Unanswered questions are whether or not there is a dilatation, whether the product is always a stack of twins and why, and what determines the operative crystallography from the limited number of possibilities.

4.4.3. Bainitic and Other Intermediate Transformations

These transformations are generally regarded as examples of semicoherent growth but with kinetic features limited by diffusion processes. As far as crystallography is concerned, they should be similar to martensites. However, insufficient data are available to state that this has been established.

4.4.4. Order-Disorder Transformations

Certain of these transformations which involve a first-order change and a change of lattice symmetry (e.g. Au-Cu) seem to have crystallographic features essentially equivalent to those of martensitic transformations. Since ordering proceeds by atomic interchange, the glissile nature of the martensitic interface is not important, but rapid growth and low strain energy both require an invariant plane at the interface.

4.4.5. Discontinuous Precipitation and Eutectoid Reactions

These transformations are nucleated mainly at grain boundaries or other singularities. Crystallography is certainly a factor in precipitation, since the α -phase in the precipitate usually has the orientation of the supersaturated solution from which it grows.

It is not known for certain whether there are significant orientation relations between the product and parent phases in eutectoidal reactions, or whether the habit planes of the plates have crystallographic significance.

4.4.6. Noncrystallographic Transformations

There are many transitions, especially at high temperatures, in which crystallography is not a significant factor. This also applies to some fast reactions such as massive transformations (Massalski, 1958).

4.4.7. Crystallography of Growth

As indicated above, the morphology of the product may be determined by either nucleation or growth considerations. We might expect growth rates to be orientation dependent, especially in interface-controlled growth. This is discussed in section 7, but there is little systematic information currently available.

4.5. Critical Questions and Suggestions for Future Work

(a) Are there any boundaries of the "epitaxial semicoherent" type?

(b) How does the transition from a coherent to an incoherent boundary occur?

(c) Are all martensitic plates stacks of parallel twins? If so, why?

(d) What is the structure of the interface when fine twins are formed?

(e) Is there really a dilatation parameter involved in martensite crystallography?

(f) Is the model of the lenticular interface in section 4.2.6 correct?

(g) Are Widmanstätten habit planes determined by energy considerations at the nucleation stage, or by growth kinetics?

(h) What are the relative surface energies of the different kinds of interfaces?

(i) In transformations with crystallographic features, it is often not known whether the observed features are determined by nucleation or by growth considerations. The problems involved are similar to those in deciding the same question in relation to recrystallization textures. Possible studies of the parameters relating to growth rates are considered in section 7; in addition, transmission electron microscopy on the early stages of many reactions should be informative.

(j) More attention might be paid to theoretical calculations of the nature of the various kinds of interfaces, and especially of the atomic arrangements. This has been valuable in recent unpublished work on mechanical twinning by Crocker, and in principle should enable the surface energy to be calculated.

(k) Experimental measurements of the surface energy of interfaces are very difficult, but some attempt might be made to carry out more systematic studies in simple systems. More accurate methods for measuring interfacial energies should be sought.

(l) The {100} precipitation of f.c.c. from f.c.c. is very challenging and a simple explanation should be found.

5. ORDER-DISORDER AND CLUSTERING REACTIONS

5.1. Order-Disorder Reactions

5.1.1. Statement of the Problem

A complete description of the equilibrium state of an ordering system as a function of temperature and composition requires more information than shown in the usual phase diagram, inasmuch as the ordering transformation occurs over a temperature range. It is necessary to know the thermodynamic nature of the ordering

transformation, since phase changes of the second kind are possible; and even when the phase change is of the first kind, important changes in thermodynamic functions occur well above and below the transformation temperature. This thermodynamic complexity affects the mechanism and kinetics of the reaction, in that both the details of formation of the ordered phase and its approach to equilibrium must be considered. For a thorough understanding of the ordering process, it is necessary that the kinetic measurements distinguish between these two effects.

Despite these difficulties, however, careful studies of order-disorder transformations can be extremely valuable, because in some cases, quite simple physical models appear to be applicable and unusually detailed comparisons between theoretical predictions and experimental results may be possible. If this class of reactions can be satisfactorily understood, the theoretical techniques may well be extended with confidence to more complex problems in solid-state transformations.

5.1.2. State of Knowledge

Despite a vast amount of work, both theoretical and experimental, our knowledge in this area is severely limited. A recent review by Guttman (1956R) covers most of the literature on the subject, so that details need not be given here.

Order-disorder reactions have been observed in a large number of systems. The crystallography of many of these transformations is known, and the transition temperatures have been reasonably well established, but not much else. Thermodynamic functions have been reliably measured for very few systems, largely because the reactions are often sluggish and equilibrium is difficult to achieve. In most cases, the kind of reaction (first or second kind) is not definitely established.

The degree of long-range order below the critical temperature has been studied in many systems, but usually by an indirect method, such as by determining the changes in lattice parameter, tetragonality, elastic properties, etc. Since the relation between such properties and the degree of order is not known with certainty, the interpretation of the results is open to some question. Accurate direct measurements based on superlattice-line intensities is possible but difficult, and very few data of this sort exist. In some cases, at-temperature measurements are necessary.

Observations of ordering kinetics have been confined largely to the Cu-Au system and to the use of indirect methods. Interpretation of the observations is clouded not only by the difficulties

mentioned above, but also by the possible simultaneous occurrence of several processes: formation and growth of ordered domains, increase of order within domains, and migration of antiphase domain boundaries. Until recently, the most successful quantitative work was that of Feder, Mooney and Nowick (1958) who studied the ordering of Cu_3Au under conditions such that only the second process (increase of order within domains) occurred. More recently, however, Pashley and Presland (1959) have been able to observe the motion of antiphase boundaries during the transformation from CuAu I to CuAu II by the technique of thin-film electron microscopy.

The theoretical situation is rather unsatisfactory, in that no exact solution for the simplest physical model, the Ising lattice, has been obtained for any three-dimensional case. A variety of approximations have been developed (Guttman, 1956R) but their reliability is unknown, inasmuch as no good estimate of the errors involved can be made.

5.1.3. Current Theoretical Ideas

On the basis of the limited data available, it seems likely that order-disorder transformations in close-packed crystals are of the first kind; those in nonclose-packed crystals are of the second kind. A plausible argument for this situation can be made (Guttman, 1956R), but no formal proof exists. The basic point is that in a close-packed structure, nearest neighbors of a given atom are nearest neighbors of each other; this is not true in the body-centered lattice, for example.

From a theoretical viewpoint, the most satisfactory systems to study are those in which ordering occurs without reduction in rotational symmetry or without any large volume change. In such instances, e.g., CuZn and Cu_3Au , strain-energy effects should be small and we may hope that the ordering process can be dealt with successfully by considering only pairwise interactions of atoms, dominated by nearest-neighbor interactions. The resulting statistical mechanical problem, although not solved, is at least clearly defined, and current work using Monte Carlo techniques (similar to those applied to other phase transformations) seems quite promising. The poor agreement observed between approximate solutions to the problem and experimental results has often been attributed to the inadequacy of the model used, but it now seems possible that the fault lies in the approximations themselves. If the model can be shown to be adequate, detailed predictions of both the equilibrium state of the system as a function

of temperature, and the kinetics of the transformation should be feasible without the introduction of any arbitrary parameters.

The nature and kinetics of an order-disorder reaction should depend on whether it is of the first or second kind. For transformations of the second kind, the curve of free energy as a function of α_1 (see sec. 3.3) for a given composition has a single minimum at each temperature. No barrier to nucleation exists, and ordering should occur without nucleation in the classical sense. The process should occur homogeneously throughout the material (see sec. 1, Table 1), with a network of antiphase boundaries determined by the vagaries of vacancy motion and the pre-existing imperfections.

For transformations of the first kind, the free-energy curve as a function of α_1 for a given composition has a double minimum at temperatures near the critical temperature, and nucleation should be required for both the ordering and disordering reactions. At temperatures sufficiently far removed from the critical, however, no barrier should exist, and ordering or disordering should proceed homogeneously, in a fashion similar to transformations of the second kind.

If the simple model is not applicable, such as when the strain energy is significant, a large energy barrier should exist, and nucleation may become quite difficult. Under these conditions, heterogeneous nucleation may well occur at preferred sites.

The kinetics observed by Feder, Mooney, and Nowick (1958) in Cu_3Au indicate that increase of order within domains occurs by the motion of vacancies at the equilibrium concentration. Under different experimental conditions, however, involving quenching from high temperatures, quenched-in vacancies should be important both in facilitating atom interchange and providing heterogeneous nucleation sites after condensation. Such effects may account for various anomalies in ordering kinetics that have been observed.

5.1.4. Critical Issues and Suggestions for Future Work

It is vital to establish the range of validity of the model in which only short-range pairwise interactions exist, that is, in which the energy is assumed to be a linear function of α 's (see sec. 3.3) with rapidly converging coefficients. If this is a good first approximation, the consequences can be worked out in detail, and the effects of nonlinearity, strain and other complications may perhaps be included as fairly simple corrections.

The roles which homogeneous nucleation and heterogeneous

nucleation play in the various transformations need to be clarified.

The existence or nonexistence of reliable functional relations between physical properties, such as resistivity and lattice parameters, and the degree of order should be established.

The following kinds of investigations may prove extremely fruitful, particularly if all carried out on the same system:

(a) Accurate determination of the equilibrium degree of long-range order as a function of temperature by careful measurement of superlattice-line intensities.

(b) Measurement of the thermodynamic functions of ordering systems in which the critical temperature is sufficiently high so that equilibrium can be established. Specific-heat measurements are especially valuable in providing very sensitive tests of theory.

(c) Measurement of ordering (and disordering) kinetics with techniques which discriminate among the types of change which occur. Possible methods include studies of superlattice-line intensity and shape, and thin-film electron microscopy.

(d) Experimental investigation of vacancy motion in ordering systems by diffusion studies and by other methods, such as vacancy quenching or nuclear magnetic resonance techniques, if possible.

(e) Theoretical analysis of vacancy motion under the conditions present in an ordering system.

5.2. Clustering Reactions

5.2.1. Statement of the Problem

Clustering may be regarded as an extremely simple case of precipitation in which the new phase has the same crystal structure as the parent phase and a similar lattice parameter; thus, lattice continuity is possible. As with ordering reactions, there appears to be some hope that the same simple model will be applicable, so that both the equilibrium structures and the kinetics involved in the clustering transformation may be analyzed in complete detail. The insight gained from understanding the cases for which the model applies (assuming there are some) should be very useful in dealing with the more complicated precipitation reactions.

5.2.2. State of Knowledge

There is a considerable body of information on clustering (Guinier, 1956R, 1959R), stemming from X-ray work on the atomic configuration of the clusters, and kinetic data obtained largely by resistivity measurements. Much of this work has been

concentrated on aluminum-base alloys, presumably because of their relation to commercially important age-hardened alloys and their attractiveness from an experimental point of view. Inasmuch as the clustered phase is metastable in most cases, care is required to avoid ambiguities due to the precipitation of more stable phases during the latter part of the clustering process. Results are mainly limited to relatively dilute solutions, since solid solubilities in the terminal phases are quite restricted in most of the systems studied.

Some thermodynamic data exist, but the problem of metastability makes both experiment and interpretation difficult. Recent kinetic measurements (Turnbull, Rosenbaum, and Treafis) provide convincing evidence that clustering in aluminum-base alloys after quenching proceeds by means of the motion of quenched-in vacancies. The rate of clustering is much lower, after reversion, presumably because a decreased concentration of vacancies is available.

Thin-film electron microscopy has recently provided much detailed information on the growth of clusters (Nicholson, Thomas and Nutting, 1959). This work indicates that in the alloys studied clusters form without heterogeneous nucleation and are competitive with ordinary precipitation. The growth of the clusters is apparently limited by kinetic effects alone. In regions where the solute has not been depleted by the competing precipitation, clusters continue to grow.

5.2.3. Current Theoretical Ideas

For the simple model of pairwise interactions, the statistical mechanical problem for clustering is essentially the same as for ordering; only the sign of the first interaction-energy term is changed. In comparison with ordering, however, clustering has received relatively little attention. Most of the theoretical analyses have been based on modifications of the quasi-chemical approach, which is inadequate for ordering and presumably also for clustering. As with ordering, the structure of the clustered system can be described by a set of α 's, and a set of free energy curves as functions of α_1 is required for a proper understanding of the transformation.

Systems in which the clustering is isotropic in shape are probably rather rare, and seem to require an almost perfect match of lattice parameters between the parent and clustered regions. The commonly observed shape-anisotropy of clustering is presumably associated with a reduction of lattice strain or with orientation-dependent elastic constants. Proper inclusion of the strain-energy

factor is necessary but difficult; the strain energy may constitute an important barrier to nucleation, and if large enough, may result in the heterogeneous nucleation of the precipitate, rather than in the formation of coherent clusters.

Most analyses of the formation and growth of clusters depend principally on the methods of classical thermodynamics and involve factors such as fine-scale surface energy, which are not amenable to direct or independent measurement. As may be true in ordering reactions, for those systems to which the simple model applies, a completely atomistic description of the transformation should be possible, using only energies measurable in separate experiments.

5.2.4. Critical Issues and Suggestions for Future Work

As with ordering reactions, it is important to determine whether systems exist to which the simple model applies, and then to investigate the effects of such complications as strain energy. The question of cluster-nucleation should be explored: under what conditions is nucleation required, does nucleation occur homogeneously or heterogeneously, and what is the nature of the heterogeneous sites?

The following kinds of investigation should help considerably:

(a) Appropriate binary systems should be selected for study. The clustered phase should be either stable or highly metastable, so that measurements of equilibrium clustering and thermodynamic functions can be conducted. Also, the clustering should be isotropic to increase the probability that a simple model may be accurately applicable. An extensive range of composition is highly desirable for these studies.

(b) Measurements should be made of thermodynamic functions, equilibrium clustered structure, and the kinetics of transformation in the above systems. The kinetics should be studied, when possible, by methods which distinguish between the formation of new clusters and the growth of old ones. Magnetic techniques may prove particularly helpful in this connection.

(c) Accurate theoretical investigation should be conducted for the consequences of the simple model, to predict the equilibrium structure, the thermodynamics, and the kinetic behavior. Monte Carlo techniques may prove useful.

(d) Direct observation of clusters should be carried out by thin-film electron microscopy, particularly with regard to the nature of the nucleation sites.

(e) Self-diffusion and vacancy studies in clustering systems should be pursued systematically.

6. NUCLEATION

6.1. Statement of the Problem

Almost all solid-state transformations occur by nucleation and growth (Bradley, 1951R; Hollomon and Turnbull, 1953R). In such a mechanism, regions in which the transformation has taken place are clearly separated from regions in which little change has occurred. The transformation proceeds by the creation, called nucleation, of regions of transformation and the subsequent growth of these transformed regions.

The essential problem before us here is that we would like to know under what conditions a particular transformation can nucleate and grow, and at what rate. Of several possible reaction mechanisms, the observed one is the one with the fastest overall transformation rate; it is not necessarily the one with the fastest nucleation rate nor with the fastest growth rate. This fact must be remembered in this, and the subsequent section where the factors involving nucleation rates and growth rates are considered individually.

6.2. State of Knowledge

The basic theory of nucleation and its experimental verification have been mostly concerned with nucleation of a new fluid (gas or liquid) phase from an original fluid phase. It is only in these instances that the surface tension of a flat interface between bulk phases has been measured and compared with the apparent surface tension obtained from a nucleation experiment. In fluid-fluid nucleation, even heterogeneous nucleation theory has been verified experimentally in part since the only additional parameters are the contact angles of the fluids on the catalyzing surface and the root angle of any scratches on this surface (if it is solid). But even in fluid-fluid nucleation theory, there is controversy regarding how well theory fits experiment. The lack of agreement, if any, between theory and experiment is thought to be due to either (a) a change in surface tension due to curvature, (b) erroneous assumptions in the derivations of the equations for the nucleation rate, or (c) unsuspected heterogeneous nucleation.

Our knowledge about nucleation from the solid is relatively meager. Almost all nucleation in the solid is found to be heterogeneously catalyzed. The type of heterogeneity responsible for

catalyzing nucleation is known in only a few instances. Even then, the quantitative aspects of how the site influences nucleation are not well understood. In addition, because the appropriate surface tensions have never been measured, we find that there is no quantitative experimental verification of nucleation from the solid even for the few observed cases of homogeneous nucleation. Thus, we know from experiment only that there generally appears to be a barrier to nucleation, but we have been unable to check it quantitatively.

6.3. Current Theoretical Ideas

6.3.1. Limits of Metastability

One of the basic controversies in the field of nucleation has been concerned with the importance of the spinodal.¹ For fluids, there is no question about its usefulness. There the spinodal marks the limit of metastability of the original phase. Beyond the spinodal, there are always ways in which the original phase can transform with a monotonic decrease in free energy even when surface tension is taken into account. Near the spinodal in the metastable region, the classical nucleation theory is expected to break down, and several theories for computing nucleation energies there exist (Cahn and Hilliard, 1959).

Both the classical nucleation-energy calculation and the spinodal concept as a limit of metastability of fluids originated with Gibbs (1948)², and he clearly states when each is applicable. In addition to the spinodal as a limit of metastability (the chemical potential of each component must increase with increasing density or composition of that component if a phase is to be stable to compositional fluctuations), Gibbs also derived other requirements of metastability for fluids, e.g. positive heat capacity and positive compressibility.

All of the foregoing pertains to fluids; the introduction of a crystal lattice complicates matters considerably. In the case of solids, if the lattice parameter changes with composition, strain energy can form an effective barrier relative to the type of composition fluctuations to which a fluid is unstable inside the spinodal. What is needed, then, is a statement of the limits of metastability for solids. It seems probable that the shear moduli must be positive, and it is likely that if one of the shear moduli tends to zero a phase transformation will be possible with a nucleation energy

¹ The spinodal is the locus of temperature-composition points at which the second derivative of the free-energy with respect to composition is zero.

² Collected Works, pp. 105-115, 252-258.

which also tends to zero, and which cannot be calculated correctly by the classical-nucleation approach.

It may be useful to regard a martensitic reaction at M_s as resulting from an unstable situation around a frozen-in defect. Some models (Bilby and Christian, 1955R) have been suggested for such a process; e.g., a stacking fault in the f.c.c. lattice is effectively a two-dimensional nucleus of a h.c.p. phase, and a "quarter-dislocation" of the f.c.c. structure gives a local atomic arrangement rather like that of a b.c.c. phase. In a martensitic transformation, the critical stage of nucleation probably corresponds to the situation in which the nucleus is just able to grow without thermal activation, rather than to the condition that an increase in size tends to decrease the free energy. Existing defects may sometimes be able to propagate into martensitic plates without thermal assistance as soon as the driving force becomes sufficiently large. For some reactions in which isothermal martensite is formed, it is known that nucleation is the slow step. Presumably the defects at which nuclei form are then unable to produce martensite until some thermally activated atomic rearrangement has taken place.

Because nucleation and growth form a two-stage selection process, the absence of a nucleation barrier does not mean that the transformation will proceed by this mechanism. For instance, if a system is unstable relative to composition fluctuations, its decomposition rate by composition change will still be governed by diffusion. If a martensitic (diffusionless) reaction under these conditions has a reasonable nucleation rate, it will be the dominant mechanism.

In nucleation theory, it is usually assumed that the original phase is at (metastable) equilibrium. However, there are cases in which the parent phase may have a nonequilibrium structure. For example, a quenched specimen may have a nonequilibrium concentration of vacancies or a nonequilibrium order parameter. Such a specimen might be unstable to continuous change and would tend to relax continuously to its (metastable) equilibrium structure while nucleation of the new phase was occurring. Two extremes may be noted. The first is the case where the time constant for nucleation is short compared to the rate of relaxation toward (metastable) equilibrium. One may then consider nucleation as occurring from the original (unstable) phase, and any changes in the original phase during the nucleation period can be neglected. At the other extreme, there is the reverse situation in which (except for a transient) nucleation occurs from the parent phase in metastable equilibrium.

6.3.2. Evaluation of the Work of Forming a Critical Nucleus

One of the most important parameters in nucleation theory is the energetic barrier to the formation of a new phase. For fluids this is computed from the thermodynamic driving force for the formation of the new phase in bulk, from the surface tension between the parent and product phases, and, in the case of heterogeneous nucleation, from the geometry of the pre-existing surfaces and the contact angles. Even for fluids, there is some doubt about the validity of using bulk thermodynamics, since the nuclei of importance are of the order of 10\AA in size. It may be better to compute the energetics of such a nucleus without the splitting into surface and volume terms. However, to evaluate properly the thermodynamic properties of such an inhomogeneity is very difficult, both experimentally and theoretically.

When we come to solids, a number of new variables appear. Even homogeneous nucleation in a perfect lattice may require vacancies to perform the atomic rearrangements. Strain energy must also be considered, and there is a choice of orientations and shapes, a choice of coherency, sometimes a choice of phases, and a choice as to whether the new phase is to be perfect. Each of these alternatives entails a separate calculation of the nucleation energy.

The structural features responsible for heterogeneous nucleation have not been characterized completely. Pre-existing surfaces, such as grain boundaries, and dislocation lines are well established sites for nucleation. For the former, the heterogeneous nucleation theory developed for fluids has been used without essential modification. For dislocations, only one simple calculation (Cahn, 1957b) for incoherent nucleation exists. This shows a large effect (compared to homogeneous nucleation) on the nucleation energy if the critical nucleus is small relative to the dislocation stress field, and little effect if the nucleus is large. The list of heterogeneities could be extensive and may include all manner of quenched-in structures. There is some value in a theoretical enumeration and evaluation of what sort of structures are apt to be important, but direct experimental studies would be far more useful.

Since the computation of the nucleation energy requires accurate thermodynamic knowledge about minute amounts of matter with highly curved surfaces, future advances will depend heavily on experimental and theoretical work in these more basic fields, i.e., inhomogeneous thermodynamics and surface thermodynamics. There are many reasons for believing that the structure or composition of a critical nucleus does not have to resemble that of the

bulk phase, and hence the initial interface need not be like the one observed between the ultimate phases. It might be expected that the nucleus will be coherent, whenever possible, even though the equilibrium phase is not coherent, mainly because strain energies vary as the volume, while dislocation and interfacial energies are much larger by comparison for the small particles.

6.3.3. Evaluation of the Nucleation Rate

At the present time, one cannot predict the nucleation rate to within several powers of ten, even in some of the best understood systems. The reason for this is that a small error in the nucleation energy corresponds to a large error in the nucleation rate. However, because the nucleation rate varies very rapidly with temperature and composition, the prediction of the temperature and composition at which nucleation will become perceptible seems more promising. For heterogeneous nucleation, calculation of the actual nucleation rate is usually not needed, because the temperature and composition interval between where nucleation at a given type of site can first be detected and where the nucleation rate is so rapid that these sites are saturated early in the reaction, is expected to be small. Hence, it is sufficient to know the density of nucleation (Cahn, 1956, 1957a). This saturation effect is expected for any nucleation catalyst whether it be a surface, a line, or a point defect. For homogeneous nucleation, the transformation rate depends on the rate of nucleation, which is more difficult to predict accurately. Nevertheless, most of the theoretical effort has been concentrated on calculations of the homogeneous nucleation rate.

The major uncertainty in the prediction of the nucleation rate has been in the evaluation of the work of forming a critical nucleus. Very little attention has been paid to the other factors in the expressions for the nucleation rate: the pre-exponential term, the diffusional term, and the general form of the expression; i.e., whether the work of forming a critical nucleus properly belongs in a Boltzmann expression. For an accurate prediction, these factors must be considered carefully, but for most systems at present, such issues are only of academic interest because of the uncertainty in determining the work of nucleation.

There is some concern about the question of an operational definition of nucleation rate (Cohen, 1958). We measure the rate of appearance of particles of detectable size. On the other hand, nucleation theory attempts to predict the rate of appearance of particles usually much smaller than those observed experimentally. For systems in which the time for the nuclei to grow to meas-

urable size is negligible, no problem exists; for other systems adequate corrections have to be invoked.

6.4. Other Aspects of Nucleation

6.4.1. Two-Stage Nucleation

In pearlite the nucleation rate seems to increase with time. This has been interpreted as the result of a two-stage nucleation process; that is, the pearlite cannot grow until both phases have nucleated and are properly positioned (Hillert, 1960). This effect may be widespread and the details should be studied.

6.4.2. Auto-catalyzed Nucleation

At times a transformation may proceed by bursts in which an original nucleation event triggers off neighboring sites. This has been observed in martensite as well as in the "group-nodule" formation of high temperature pearlite. In considering the overall transformation kinetics, it is more convenient to treat this type of nucleation as a part of the growth rate, but full understanding of the details must await more basic knowledge concerning the nucleation process itself.

6.4.3. Two-Dimensional Nucleation as a Growth Mechanism

The finding of whisker morphologies for precipitates suggests that crystal growth from a solid may sometimes proceed in a manner similar to that from a fluid. In this case, thickening of whiskers may depend on two-dimensional nucleation of new crystal planes.

6.5. Suggestions for Future Work

(a) No methods currently exist by which direct measurements can be conducted on that transient creature, the critical nucleus. Very few measurements of the nucleation rate in solids have been made, and most of these are of little basic value because it is usually not known whether the nucleation is homogeneous or heterogeneous, and, if heterogeneous, the nature of the nucleation site has not been identified. The most powerful technique for studying nucleation is the small droplet technique developed by Turnbull and Vonnegut and applied to solids in only two instances: Cech and Turnbull (1956) for martensite, and DeSorbo and Turnbull (1956) for Pb-Sn. The basis of this technique is that in sufficiently small particles, heterogeneous catalysts may be absent from most of the particles, and only one such catalyst may be

present in the others. The technique therefore affords an opportunity to study both homogeneous nucleation and heterogeneous nucleation. It should be possible to isolate and characterize the individual heterogeneities, that is, to determine whether a broad spectrum of activity exists for a given type of catalyst or whether each type of catalyst has a discrete catalyzing power. It should also be possible to relate a particular type of nucleation behavior to a particular type of site.

(b) Whereas it may be almost impossible to measure the work of forming a critical nucleus, it should be possible to track down the sites responsible for nucleation, using transmission electron microscopy.

(c) A possible way of finding and studying the sites responsible for nucleation is to investigate the effects of willfully-created defects, such as vacancies produced by electron bombardment, dislocations produced by plastic deformation, polygonized boundaries produced by deformation and recovery, etc. Ionic and semiconductor materials may be particularly useful for these studies.

(d) Nucleation theory depends on the thermodynamics of extremely small particles which cannot be stabilized for thermodynamic measurements. It is, however, possible in a number of systems to form fully precipitated particles in the appropriate size range for equilibrium. Thermodynamic measurements should be made in such cases since much can be learned that is directly applicable to nucleation theory. Both the nucleation rate and the thermodynamics should be studied in the same system.

(e) Inasmuch as the choice of a mode of nucleation in a given reaction almost certainly depends upon competition among various mechanisms, some system in which several possibilities compete should be investigated to find whether they proceed independently until impingement or whether they catalyze or inhibit one another.

(f) In order to obtain a clear picture of the effect of imperfections on the nucleation of martensitic transformations, it would be desirable to have additional kinetic data on transformations which proceed isothermally in the absence of prior athermal martensite. Furthermore, it would be especially valuable to make such observations on alloys exhibiting isothermal martensite *above* M_s .

7. THE GROWTH PROCESS

7.1. Statement of the Problem

In general, a growth process may involve a number of consecutive steps, each of which can take place by several alternate mechanisms. The problem is to understand the basic nature of these

individual steps as well as their interaction with one another. The approach adopted here is indicated by the following outline: section 7.2 deals with the interface movements by which atoms of the parent lattice rearrange themselves to generate the new lattice. Section 7.3 treats the long-range diffusion by which the compositional changes of the phase transformation are acquired. Section 7.4 discusses some of the complicating factors which influence the details of phase transformations. The interplay among the various steps is analyzed in section 7.5.

7.2. Current Knowledge and Ideas Concerning Boundary Movements

This section discusses the atomic processes near an interface which are necessary for the motion of the interface. As described in section 4, the boundary may be coherent, semicoherent or incoherent, and it may be sharp or diffuse. Two types of motion are discussed. A glissile interface is defined as a boundary which can move without thermal activation, and is thus mobile even at temperatures near 0°K. The motion of all other interfaces is thermally activated, and must become negligibly slow at sufficiently low temperatures.

7.2.1. Coherent Glissile Interfaces

A glissile boundary will move when subjected to either chemical or mechanical driving stresses, and may be fully coherent or semicoherent. The mechanism of growth is believed to differ in the two cases.

Experimental work on the transformation in cobalt and cobalt-nickel alloys seems to demonstrate that coherent growth cannot be described in terms of the independent movements of individual atoms or of successive atom planes. If this were so, the direction of motion of the atoms would have a component in the habit plane which changed from one to another of the three possible $\langle 112 \rangle$ or $\langle 1\bar{1}00 \rangle$ directions every few atom planes, thus avoiding a shape deformation, and its associated coherency stresses, almost completely. There is no doubt, however, that in cobalt-nickel alloys some thousands of (alternate) $\{111\}$ planes move over each other in the same $\langle 112 \rangle$ direction during transformation from the cubic to the hexagonal phase. In pure cobalt the position is less certain, but regions of homogeneous shape deformation have been found in single crystals. Transformation under mechanical stress gives rise to much larger regions of

homogeneous shape deformation, especially in alloys. (Takeuchi and Honma, 1957; Gaunt and Christian, 1959).

It seems improbable that a fully coherent twin boundary, or martensitic interphase boundary, (without atomic steps) could be mobile under applied stress. The extension of one grain by one unit cell is more likely to begin with the formation of a small island or two-dimensional nucleus, surrounded by a closed step of atomic height. This step is a twinning or transformation dislocation, which can glide in the boundary surface and expand under applied stress, but in the absence of a sufficient stress, it will collapse again because of its line tension. In this idealized model, there is thus a nucleation difficulty on each successive lattice plane. The required steps might be nucleated at regions of internal stress, or at places where the coherent boundary ends on a grain boundary or a free surface.

The theory of growth of a plate with a coherent interface is largely due to Bilby (1953) and has been applied to cobalt by Seeger (1953, 1956). The description is essentially the same as that used in the Cottrell-Bilby theory of mechanical twinning, and depends on the topological properties of suitable dislocation nodes, formed when lattice dislocations pass through the boundary. The transformation dislocation is thereby enabled to glide on a helicoidal surface, moving the interface forward through successive lattice planes as it does so. Although there is no direct evidence for this proposal, it appears to be the only convincing mechanism for the motion of a coherent interface. The experimental evidence for cobalt-nickel alloys, and less certainly for pure cobalt, precludes the possibility of independent nucleation of steps on successive planes.

7.2.2. Glissile Semicoherent Interfaces: Fast and Slow Growth

Martensitic transformations may be divided into two categories: those in which the growth is slow enough to be observable under a microscope, and those in which individual plates are fully formed within a small fraction of a second. The growth velocity in the second class is probably an appreciable fraction of the velocity of sound, but has been measured only in an iron-nickel-carbon alloy (Bunshah and Mehl, 1953). Slow *isothermal* growth has been observed in the transformation from β to α uranium in a dilute uranium-chromium alloy, but in most other transformations, slow growth seems to require a changing temperature. In general, slowly moving interfaces are encountered in transformations with small accompanying shape deformations, and such interface

motion begins at comparatively small values of the chemical driving force. Some slowly moving plates have been shown to be in thermo-elastic equilibrium with the matrix, growing or shrinking as either the temperature or an external mechanical stress is changed. This behavior is analogous to elastic twinning.

In thermo-elastic martensite, the rate of growth is determined entirely by the rate at which an external constraint can be varied. With the usual approximations of elastic theory, the driving force and the accommodation strain energy are both proportional to the volume of the plate, provided its shape remains constant. Ultimately, the edgewise growth will be constrained, and as the plate thickens, the strain energy rises much more rapidly than the chemical free energy decreases. If a minimum of energy is attained without appreciable plastic deformation of the matrix, there is equilibrium between the driving stress and the opposing stress, and the interface may be made to move in either direction by slight variation of the external conditions. Thermo-elastic martensite formation is thus favored by small shape deformations and a high yield strength of the matrix.

This interpretation of fast and slow growth fits most of the data, and implies that important kinetic differences in martensitic growth do not arise from fundamental differences in the nature of the interfaces. The alternative description, used for example by Suzuki (1954) and Crussard (1955) is that there are two different types of interface and growth mechanisms. This is presumably true for the slow isothermal growth in the uranium-chromium alloy, which requires an unusually large shuffling of the atoms within the unit cell. It is possible that this slow growth is connected with the presence of atoms on special sites, since the transformation in pure uranium is much more rapid.

The dislocation model of a semicoherent martensitic interface insures either that all the dislocations are parallel and can glide readily as the interface moves, or that the dislocations present are equivalent to a single glissile set. Bullough and Bilby (1956) have shown that the assumption that the interface is a simple glissile surface dislocation leads to crystallographic theories which assume that the interface is a plane of zero average distortion. The mobility of the dislocations should be sufficient to insure that the whole interface is glissile, and there is no need to invoke a Cottrell-Bilby pole mechanism. This is a property of the dislocation array giving the martensitic crystallography; other types of semicoherent interface will not be glissile.

If a martensitic plate is a stack of fine twins, the twin boun-

daries become the physical discontinuities at the interface. When large twins are formed, motion of the boundary must be accompanied by the movement of arrays of twinning dislocations behind the boundary, or possibly of the slip dislocations ahead of it. There is experimental evidence that the twinning dislocations move less rapidly than the interface itself, and it may be significant that large twins are formed in gold-cadmium, indium-thallium and copper-manganese alloys, all of which show relatively slow growth.

The above discussion is concerned with the motion of the habit plane interface normal to itself. A lenticular plate could extend parallel to the habit plane by expansion of transformation dislocation steps (sec. 4.2.3). In the direction in the habit plane parallel to the slip dislocations or twin boundaries, this process requires simultaneous expansion of the dislocation or twin boundary loops, and should be readily accomplished. In the direction in the habit plane normal to this, new dislocation loops or twins must be nucleated as the crystal grows; this could be a relatively easy process because of high stress concentrations near the tip of the plate.

7.2.3. Single Interface Transformations

Slow growth is also observed in single interface transformations in gold-cadmium and indium-thallium alloys, where the possibility of thermo-elastic equilibrium is precluded. The interface moves in a jerky fashion, and the temperature must be continually lowered or raised to maintain its motion. There is considerable evidence that the interface may move rapidly between obstacles, and audible clicks are heard in some circumstances.

These observations seem to imply the existence of an opposing stress which increases with the volume transformed or with the distance moved by the interface. However, there is some possibility, particularly in indium-thallium alloys, that the observed behavior may be due to the combined effects of slight composition gradients in the original crystal and temperature gradients in the experimental arrangement. Some support for the existence of an opposing volume stress ("volume relaxation") comes from the observations of Bainbridge, Li and Edwards (1954) on the stress-induced migration of a low-angle grain boundary. At low temperatures, an increasing stress is required to maintain this motion; this result is attributed to the sweeping up of dislocations, etc. by the moving boundary. There is also a possible analogy with the pseudo-elastic behavior of twin boundaries at low temperatures.

7.2.4. Cessation of Growth: Stabilization

There is little systematic information on the effects of substructure, impurity particles, etc. on the growth of martensitic plates. Edgewise growth often continues until stopped by grain boundaries or other plates, and smaller plates are produced in the late stages of transformation because of partitioning of the matrix.

Martensitic plates which form very rapidly do not usually increase in size on further lowering of the temperature; they may be called stabilized plates, in contrast to the thermo-elastic plates which are in equilibrium with the matrix. For thermo-elastic plates formed on cooling, the motion may be reversed on reheating, and the whole transformation is crystallographically reversible. This behavior is also characteristic of some transformations (e.g. polycrystalline indium-thallium or copper-manganese) in which parallel-sided plates are produced. Complete transformation on cooling is possible and growth of a given plate does not stop until it impinges on a parallel plate; in a macroscopic sense, therefore, the only boundaries remaining after the cooling transformation are those between adjacent plates of product. The observed reversibility implies that nuclei are retained or reformed at these interfaces, which remain glissile.

A stabilized plate disappears only when heated to a temperature considerably above M_s , and the transformation is not completely reversible inasmuch as surface tilts often remain. (The plates may or may not transform to the original grain structure of the parent phase: such crystallographic reversibility is often a consequence only of the lower symmetry of the product phase). In zirconium, the plates seem to disappear reversibly on heating if the pre-cooling is carried only slightly below M_s , but permanent surface tilts are produced if the specimen is cooled a few hundred degrees below M_s before reheating. The surface tilts produced by the cooling transformation in cobalt-nickel alloys are not removed by the reverse transformation.

Holding at constant temperature below M_s frequently stabilizes the parent phase, transformation on subsequent cooling being retarded. In transformations where plates grow to full size immediately, stabilization must involve the inhibition of either the formation or early growth stages of a nucleus. In thermo-elastic martensites, there may be some stabilization of the interface itself, and it is known that this can occur in single interface transformations. Sufficiently long holding stabilizes an interface indefinitely, transformation on subsequent cooling eventually restarting from a different nucleus. After moderate holding, the original

interface will resume the motion, but will be halted again in the same place on the reverse heating transformation. During thermal stabilization, the surface tilts become rounded in indium-thallium.

Stabilization at temperatures above M_s may be induced by mechanical deformation. Light mechanical working sometimes raises M_s , presumably by creating nucleation sites. Heavier working often inhibits reaction, as also does thermal cycling through the transformation. This effect is pronounced in cobalt, where the f.c.c. form may be stabilized to liquid nitrogen temperature by such treatment.

It seems possible that a martensitic plate may cease to grow into a matrix of higher free energy for any of the following reasons:

- (1) The interface has lost its glissile character, and can no longer move without thermal activation. This could happen because of plastic deformation in the matrix or product, or because the interface accumulates defects as it sweeps through the matrix, and these eventually jam its motion. If the Bowles-Mackenzie (1954,1957) theory of the crystallography is correct, it may also be that nonglissile dislocations are nucleated in the interface when the plate reaches a certain size.

- (2) Although still glissile in principle, the interface can no longer be forced through the matrix, which has been hardened or distorted during transformation.

- (3) The interface has been stabilized by the migration of solute atoms to its vicinity, so as to reduce the net local driving force to a low value.

- (4) Dislocations in the matrix have been pinned by solute atoms, so that the plastic deformation required to accommodate the shape change is no longer possible.

Explanations (1) and (2) will account for the normal stopping of stable martensitic plates, and for mechanical stabilization; (3) and (4) are alternative explanations of thermal stabilization of the interface. There is experimental evidence to favor (3) in single interface transformations.

7.2.5. Thermally Activated Motion: Factors Affecting Mobility

Much of our present knowledge of thermally activated processes at or near an interface during its migration is derived from experiments on grain boundaries rather than interphase boundaries. Although the driving forces in recrystallization or in grain growth are different from those in a phase transformation, the atomic mechanisms may be similar. The advantage of studying grain-

boundary migration, or phase-boundary migration in polymorphic changes, is that undoubtedly interface processes are then rate-determining.

A difficulty inherent in most such experimental work is the differentiation of the effects of varying driving force from those of varying mobility. The available evidence suggests that most low-angle boundaries of small misorientation have lower mobilities than high-angle boundaries, but that some special low-angle boundaries have very high mobility. There is a natural tendency to identify the very mobile boundaries, which are observed during polygonization or subgrain growth, with glissile arrays of dislocations, and this is supported by the effect of stress on mobility (Beck, 1954R). By analogy, it would follow that semicoherent interphase boundaries of the epitaxial kind have lower mobility in general than incoherent boundaries. However, there is no evidence for this.

In addition to the general distinction between semicoherent and incoherent boundaries, there is reason to expect certain special orientation relations to have high mobility, in that the regular incoherent boundaries are retarded in the presence of impurities. It is now believed that this accounts for the development of most annealing textures during recrystallization, and there is good evidence that in at least one precipitating system (Al-Si), the preferred orientation of the precipitate is due to a preferred growth rate rather than to a more rapid nucleation of this orientation. However, it is clear that in some very rapid transformations of the massive type, incoherent boundaries may be very mobile, irrespective of the orientation of the grains into which they migrate; the product crystals frequently grow across the original grain boundaries (Massalski, 1958).

Under constant driving force, the rate of boundary migration is constant when controlled by interface processes. The temperature variation of this velocity has been measured for grain-boundary movements and for interphase-boundary movements in transitions such as pure tin. The results are usually analyzed by use of chemical reaction rate theory, but do not always allow a single activation enthalpy to be deduced. When this is possible, the enthalpy derived sometimes corresponds to that for grain-boundary diffusion, but is frequently much higher.

Other factors may have considerable influence on boundary mobility, but systematic studies on possible effects of stress fields, perfection of surrounding lattice, etc. are not available. The interaction of various factors is difficult to separate experimentally; it

is not known, for example, in what circumstances (if any) the velocity is proportional to the driving force.

7.2.6. Theory of Thermally Activated Growth

Epitaxial semicoherent boundaries contain dislocations of the edge-type with Burgers vectors in the boundary (sec. 4.2.3). As the boundaries migrate, the dislocations effectively climb into one or the other lattice by addition or removal of vacancies or interstitials. So long as the flow of vacancies or interstitials can be maintained, it appears that the main difficulty in this motion will be the addition or removal of atoms to the extra half planes at the interface. It is not clear whether or not the dislocations will need to maintain a density of jogs (steps in the interface) to allow the motion, as would be assumed for isolated edge dislocations. Accordingly, the activation energy for the growth process may either be that for jog formation or that for lattice diffusion of vacancies or atoms.

For incoherent boundaries, the usual assumption is that the important growth process is the thermally activated transfer of single atoms across the interface. Chemical reaction rate theory (or steady-state thermodynamics) then gives a migration rate which is proportional to the driving force (when this is small in comparison with kT), and to a Boltzmann-type factor involving the activation energy for crossing the interface. This energy should be approximately that of grain-boundary diffusion, but is frequently much higher. Suggestions have been made that the high value arises because the unit activation process involves several atoms or because there is a temperature-dependent concentration of impurity atoms at the interface. Results for tin have also been interpreted by assuming that atoms can transfer at only a small fraction of active sites, this fraction being a function of temperature.

The above models assume there is no nucleation difficulty in transferring atoms from one phase to the other. In some circumstances, however, the boundary may be a rational habit plane, and if no dislocations thread through it, it is possible that a two-dimensional nucleation process will be required for growth. There is no definite evidence for this, except for the existence of some morphologies which do not seem to grow in certain directions. It will probably be necessary to invoke such a process if it is established that the growth rate in a transformation involving long-range transport or solute atoms is nevertheless interface controlled.

7.2.7. Critical Questions Concerning Boundary Movements

(a) Is a dislocation pole mechanism essential for growth of coherent twins and coherent martensite, or is there some other mechanism?

(b) Is there any fundamental difference between the interface or its motion for fast growth and that for slow growth?

(c) Is the glissile surface description of a semicoherent martensite interface correct? Are there any such interfaces with dislocations, as opposed to fine twins?

(d) What forces oppose the motion of a glissile interface? In a stabilized martensitic plate, is an interface halted because it has lost its inherent mobility or because a work-hardening or similar effect prevents its being moved?

(e) Does it really become more difficult to move a single interface into a uniform matrix at constant temperature?

(f) Are both edgewise and normal growth rates always fast except when the plates are in equilibrium with the matrix? Are there any differences in the (fast) growth rates normal and parallel to the interface? When growth is parallel to the habit plane, is there any difference in rate for the direction in which new dislocations or twins have to be nucleated?

(g) What obstacles other than grain boundaries stop edgewise growth?

(h) Is the growth in the uranium transformation thermally activated because of adjustments at the interface, because of thermally activated plastic flow in the matrix or because of diffusion ahead of the interface? Why is the transformation more rapid in high-purity uranium?

(i) Is an epitaxial semicoherent boundary less mobile than an incoherent boundary? Are there any special boundaries which are unusually mobile? Do the very mobile boundaries operative in massive transformations have any special properties?

(j) How does the rate of boundary migration depend on variables such as driving force, impurity control, temperature, structural condition, etc.?

(k) Is two-dimensional nucleation required for any solid-state growth process?

(l) What is the atomic process responsible for growth of semicoherent and incoherent crystals?

7.3. Current Knowledge and Ideas Concerning Long-Range Atom Transport

In those phase changes which involve compositional changes,

atom transport must occur. This section is concerned with how fast this transport can take place. Section 7.5 treats the question as to whether the atom transport will be the rate-controlling factor, or how the system chooses a particular geometry for the phase change and thereby influences the amount of transport required.

Most of this section falls in the field of diffusion which is treated specifically in Part V. However, there are several points that relate directly to solid-state phase transformations, and these will be considered here.

7.3.1. Diffusive Transport Through Bulk Phases

Bulk transport is governed by the diffusion equation, and depends locally on the diffusion coefficient and the activity gradients of each component, which in turn depend on composition and stress. In many systems, the diffusion coefficient also depends on thermal history through the effect of quenched-in vacancies. The length of time before the excess quenched-in vacancies disappear at various sites is of considerable importance in the kinetics of low-temperature reactions. It has been suggested, for instance, that vacancies in Al-Cu alloys are attracted to the clustered regions and are stabilized there, and that the clusters, together with their associated vacancies, move through the lattice collecting additional atoms instead of having the atoms diffuse individually through the matrix. At sufficiently high supersaturations of vacancies, it has been noticed that vacancies can disappear rapidly by another process; namely, condensation of vacancies to form voids or dislocation loops. On the other hand, such formations, if sufficiently small, will give rise to equilibrium vacancy concentrations higher than normally expected and thus enhance the diffusion rate for a considerable length of time.

7.3.2. Diffusion Short Circuits

A number of short-circuiting diffusion paths may exist; e.g., grain boundaries and dislocation lines. As in diffusion studies, these are not important in transformations at high temperatures but become increasingly important at low temperatures. These paths and the migration of quenched-in vacancies are thought to be the main mechanisms for long-range transport of substitutional atoms at low temperatures.

7.3.3. Mathematical Solution of the Diffusion Equation

It should be possible to solve the diffusion equation, either analytically or by machine methods, if the position and composi-

tion of the various interfaces (as well as the stresses exerted, since they influence the diffusion rate) are known as a function of time. At present, analytical solutions exist for a number of simple situations.

The time-dependent geometry of the various interfaces is usually fed into the diffusion equation as a boundary condition. In a number of cases in which the diffusion equation has been solved, several acceptable mathematical solutions, assuming different time-dependent geometries, have been found. However, certain indeterminacies remain to be settled. Ham (1958), for instance, has shown that *any* ellipsoid growing under diffusion control will preserve its shape. What then determines the initial shape? Moreover, Ham's solution indicates that any ellipsoid will grow with dimensions proportional to $t^{1/2}$, a result which is in striking contrast to an earlier solution found by Zener assuming a constant growth rate. What then determines the growth law for a certain reaction?

In many systems, it is likely that the growth is not solely diffusion-controlled but, to some degree, interface controlled, and this might remove the indeterminacy. Unfortunately, a theoretical treatment seems very difficult. In other cases, the indeterminacy may be removed by considering the stability of shape against fluctuations. In any growth process, the conditions are bound to fluctuate, and this gives rise to the possibility that certain geometries, while satisfying the diffusion equation and all boundary conditions, may not prevail because they are unstable with respect to fluctuations. In other words, there are factors which cause the fluctuations to increase in magnitude once they have occurred. It has, for example, been proposed that this instability is responsible under certain conditions for the appearance of dendrites instead of plane growth fronts or equiaxed particles. The test of stability would be most useful but has not yet been devised. However, Zener (1946) tentatively suggested that the shape giving the highest growth rate may be the only stable shape.

7.4. Other Factors in the Growth Process

There are many complicating factors in the growth process of solid-state transformations. Often these factors may be so important that the main features of the growth process cannot be understood without considering the complications. The following list is not intended to be complete, but rather to illustrate the kinds of difficulties that must be overcome before a deeper under-

standing of growth processes can be achieved. The present knowledge on these factors is quite limited.

7.4.1. Role of Impurity or Alloying Elements

The system may contain an element whose presence is not essential to make the reaction thermodynamically possible. Such an element may affect the growth process in many ways, most of which have not been studied closely as yet. The interaction of impurities with crystal interfaces and imperfections, and the related effects on grain boundary mobility, have recently become the subject of great interest and are discussed in some detail in Part VII, Growth, Structure, and Morphology of Crystals.

In general, it can be said that the effect of an impurity depends strongly on how it distributes itself between the parent and product phases. Although this redistribution may not be thermodynamically necessary, it may take place nevertheless and the thermodynamic conditions for the growth process may thereby change considerably. The concepts of ortho- and para-equilibrium introduced by Hultgren (1953) to describe the cases of complete and no redistribution may prove valuable.

7.4.2. Role of Coherency Stresses

Coherent precipitation gives rise to stresses which are necessary for the mutual accommodation of the parent and product lattices. The matching does not have to be perfect, however, and the misfit can then be represented by a dislocation array. The effect of such coherency stresses and dislocation arrays on the growth process is not well understood. At a critical size, the growing particle is believed to "break away" from the matrix, resulting in the formation of an incoherent interface. This transition is probably caused by the coherency stresses which increase as long as the particle grows coherently. The mechanism of this transition and its effect on the growth rate are not known. A further possibility which has not been ruled out is that the incoherent particles may be independently nucleated and dominate over the coherent particles after the latter have reached a certain size.

7.4.3. Role of Mechanical Stress and Pressure

It has been shown that a precipitated particle having a larger volume than the corresponding hole in the parent phase can minimize the elastic energy by attaining the shape of a plate, provided that the necessary volume requirement is met by elastic deformation only (Nabarro, 1940). The influence of the elastic stress field

on the growth rate and on the actual development of the shape has not been worked out in detail, however, nor has the interplay between elastic energy and interfacial energy during growth.

It has been suggested that plastic deformation may sometimes provide the extra space required for continued growth. In this instance, the elastic stress field will have a hydrostatic component whose effect on the growth process is, in principle, accessible to thermodynamic treatment although, except for the simplest cases, the calculations may be very complex.

Necessary space for a growing particle may partly be supplied by the diffusion process itself if there is a strong Kirkendall effect in the right direction or if there is a large concentration of quenched-in vacancies. The growth process may thus be intimately related to the diffusion process.

7.4.4. Role of Interfacial Energy

If the interfacial energy alone determined the shape of a particle, the transformation product would assume the particular shape which minimizes the total interfacial energy. However, the shape is usually determined by the growth process, and the interfacial energy affects the shape only indirectly through its influence on the growth process. When the growth is diffusion-controlled, the main role of the interfacial energy lies in locally changing the equilibrium composition at the curved interfaces and thus causing local variations of the driving force for diffusion. To solve the diffusion equation seems quite difficult if this dependence of the composition on the shape of the interface is to be taken into account. Only a few cases involving steady-state conditions have been treated so far, and those, only in an approximate manner. It seems that a mathematical test for the stability of a certain growth shape, mentioned in section 7.3.3, would be of value only if the influence of the surface energy is taken into account.

7.5. Current Knowledge and Ideas Concerning the Kinetics of Growth

Growth involves many individual steps or rate processes. The interface must move, the lattice may have to change, diffusion and plastic deformation may be required. Each of these processes can occur in a number of ways. What then determines the rate of growth? Is it possible to ascertain experimentally the rate-controlling step, or to predict it theoretically?

7.5.1. The Rate-Controlling Step

So far there does not seem to be any general procedure by which the rate-controlling step can be identified from experimental observations on the morphology and kinetics of growth. The customary procedure is to calculate the rate and geometry of growth by making some assumption about the nature of the rate-controlling mechanism, and then to compare the result with experimental observations. In order to carry out such calculations, the rate of all possible mechanisms must be estimated, assuming that each is governed by the same total driving force. Some of the basic mechanisms are known well enough to permit such a calculation, but others are not. In many instances, however, most of the consecutive steps can be ruled out immediately because they are much too fast to be rate-controlling.

Martensitic, massive, or allotropic transformations do not involve any long-range diffusion; thus, for many of these reactions, the growth rate may be controlled by the mobility of the interface. However, it is not clear why the system chooses a particular type of interface. This may sometimes be a result of competition, the fastest growth process transforming most of the specimen. It is also possible that the fastest nucleation process leads to a certain kind of interface. A further alternative is that one kind of interface can be transformed into another as the growth proceeds. In such a case, it may not be quite certain that the fastest moving interface is formed. Too little is known about the mechanism of interfacial movement to decide this. Furthermore, it should be emphasized that the pressure or stress created by the transformation itself may be of key importance in decreasing or completely eliminating the driving force for interface movement if the reaction advances too rapidly for relaxation to take place.

It may be quite complicated to analyze a growth process if many interrelated steps are involved. An estimate of the growth rate based on the assumption that one mechanism is rate-controlling may not be realistic or even possible to carry out. It is then valuable to carry out measurements under different conditions and to study how the growth responds to different variables. Temperature and composition might be the two most significant variables in this respect, but other variables should also be introduced, if possible, such as impurities or very high hydrostatic pressures.

There may be cases in which a certain phenomenon occurring during the growth process is not necessary in order for the growth to proceed. If a phenomenon of this type is too slow, it may then

not take place at all or take place only to a certain degree. The diffusion of carbon away from a growing Widmanstätten plate of ferrite during the formation of bainite may be such a case. Another example may be segregation during discontinuous precipitation. It is not known what determines the extent of such "unessential" processes. One might assume that these processes occur to the degree necessary for the overall growth rate to be a maximum, but this assumption has not been justified on theoretical grounds.

7.5.2. Lamellar Growth in Precipitation and Eutectoid Decomposition

Many solid-state reactions involving segregation, such as precipitation and eutectoid decomposition, result in a regular arrangement of the two final phases. A common transformation product consists of a lamellar structure in which parallel plates of the final phases are grouped in nodules or cells. Most of the segregation occurs as the boundary of the growing nodules advances through the matrix, leaving behind the lamellar structure. The purpose of study in this field is to develop an understanding of the conditions under which this morphology is to be expected and the factors that influence the growth characteristics.

In a number of systems, the rate of lamellar growth is known as a function of temperature and composition, and in a few instances, as a function of the relative orientation of the phases as well. Measurements have been made of the spacing between plates or rods of each phase. Some data are available on their composition. In recent years, it has become apparent that, at least in some cases, the rapid mass transport required for the observed redistribution of composition cannot occur by volume diffusion alone, and that diffusion along the moving nodule boundary is perhaps the only way this segregation can take place at the observed rates. However, it has been suggested (Darken et al., 1957) that in Fe-C pearlite, all the phases are not in contact with the cell boundary. This situation requires the occurrence of volume diffusion.

Growth anisotropy of a lamellar nodule may result from the inability of the reaction product to grow into both grains from the grain boundary. In many precipitation reactions, this is understandable because the lamellae of the parent phase have the same orientation as one of the grains, and hence will lack the interface necessary for rapid diffusion. In some eutectoid systems, it seems that a proeutectoid film of one of the new phases isolates the grains from one another.

Sideways growth is another cause of growth anisotropy. However, most nodules are approximately equiaxed. It could be that the plates are everywhere nearly normal to the cell boundary. If they are, they will not usually appear perpendicular on a random section.

It has recently been suggested (Cahn, 1959) that the rate of a geometrically complicated process, such as lamellar growth, requires the specification of at least two rate-controlling parameters, e.g. diffusion and cell-boundary mobility. Additional physical principles also seems called for before the essential parameters of the process can be fully specified (Zener, 1946; Cahn, 1959; Kirkaldy, to be published).

With regard to impurities, two basic suggestions for the role of alloys in steels have been offered. One proposes that the effects of alloying additions can be interpreted through their influence on diffusion and constitution; the other holds that they change the mobility of the lamellar boundary.

7.5.3. Bainitic Growth

The structure now called bainite was originally found in iron-carbon alloys transformed in a certain range of temperatures. Similar structures have later been found in other eutectoid systems. Bainitic transformations in alloys other than steel have not been studied extensively, and the following discussion will, therefore, deal mainly with the bainitic reaction in iron-carbon alloys.

Bainite in iron-carbon alloys consists of ferrite and carbide. The morphology strongly suggests that ferrite is the primary phase in the formation of bainite. As a consequence, the characteristics of bainite depend mainly on the manner in which the ferritic constituent forms, and attention should thus be focused on the mechanism of formation of the bainitic ferrite. However, the recent observation that Widmanstätten cementite may precede the first appearance of bainitic ferrite in highly hypereutectoid steels (Jellinghaus, et al., 1954) is not surprising and suggests that carbide may become the leading phase in bainite in very high carbon alloys.

Each unit of bainite consists of several plates of ferrite, the number of plates depending strongly on the temperature of formation and on the nature of the nucleation site. Each plate grows edgewise and sidewise. Sooner or later particles of carbide form in contact with the austenite on the sides of the ferritic plates and the sidewise growth is then accelerated until impingement with the neighboring plates occurs. There are reports (Speich and

Cohen, 1960) that at low temperature, the carbide forms inside the ferrite. A change in kinetics has also been found at temperatures below 300–350°C.

Kinetic measurements show that bainite forms at two widely different rates, the initial stage of the reaction being quite rapid, and the second stage considerably slower. Above a certain temperature, called B_1 , the formation and growth of bainite may stop before the reaction is completed. Another temperature, B_c , exists above which there is no formation of bainite. These two temperatures are analogous to the M_s and M_f for martensite. The observation that the formation of bainite gives rise to a surface relief effect, as does the formation of martensite, points to a similarity between the two types of transformation. Widmanstätten ferrite likewise produces surface tilts, and may be regarded as a member of the bainitic series. In fact, there is a further similarity in the ferrite-austenite lattice relationships.

Only in a few cases have detailed models been worked out for the mechanism of the bainitic reaction. In most of these instances, the similarities with martensite have been stressed, and the assumption has been made that the bainitic ferrite inherits the carbon content of the parent austenite, as is the case for martensite, and that carbon diffusion and/or precipitation occur as subsequent processes. Speich and Cohen (1960) have presented an alternate model in which carbon diffusion is an integral part of the transformation. The growth rate appears to be controlled by diffusion of carbon, but the bainitic ferrite presumably forms by a shear transformation and is semicoherent with the parent austenite. This suggestion is strongly supported by the observed orientation relationships and the surface-relief effects. The slow diffusion-controlled growth does not necessarily imply that the carbon content of bainitic ferrite is exactly that required by thermodynamic equilibrium with austenite.

An important question arises concerning whether there exists in fact a close relationship between bainite and martensite. If the two reactions are closely related, a gradual transition from bainite to martensite could be expected as the temperature is lowered. The existence of such a transition should be revealed by an increase of the supersaturation and growth rate of bainitic ferrite at sufficiently low temperatures. If this does not happen, bainitic transformations may be found in eutectoid systems, even though martensitic reactions do not occur.

An explanation for the domination of bainite over pearlite at low temperatures may be fundamental to an understanding of

bainitic growth. This problem has frequently been discussed but not yet satisfactorily resolved although several explanations have been proposed, such as different mobilities of semicoherent and incoherent interfaces, favored diffusion at sharp edges of growing particles, and favored nucleation of coherent nuclei at low temperature.

Another question that should be answered concerns the factors that determine the platelike shape of the ferritic units in bainite. The explanation may include the effect of an anisotropic interfacial energy and of strain energy during growth as well as some factor inherent in the mechanism of semicoherent growth itself.

So far we have mainly dealt with the growth of a single plate of ferrite. One should also inquire about the mechanism of formation of all the plates in a unit of bainite. For instance, the several plates may be branches from a common grain-boundary precipitate; they may be the result of random nucleation at an austenite/austenite grain boundary; or the first plate may induce the nucleation of the next one either by the stress field it causes in the parent austenite or by forming branches after having grown sidewise.

7.6. Suggestions for Future Work

7.6.1. On Boundary Movements

(a) Thin-film electron microscopy should give considerable information about the nature of the various interfaces and the way in which they move. This probably constitutes the most promising experimental approach at present.

(b) More attention should be devoted to the motion of individual boundaries. In martensitic single-interface transformations, it would be valuable to study the effect (if any) of zone levelling of the specimen and of strictly isothermal conditions during the transformations. Single-interface transformations in thermally activated processes should also be possible (they have been obtained in recrystallization) and studies of boundary mobility as a function of misorientation, driving force, impurity content and temperature, would be very informative.

(c) It would be highly desirable to obtain more data on the migration rates of fast interfaces.

(d) A study of the effect of crystal perfection on the size and shape of stabilized martensitic plates would be useful. Several workers have tried without success to obtain single-interface transformations in (say) iron-nickel alloys, but this might profit-

ably be pursued with floating-zone techniques for crystal preparation.

(e) Theoretical work on crystallography has been mainly phenomenological on martensitic transformation and has been very meager on other types of transformations. There now seem to be compelling reasons for investigating detailed atomic models, paying special attention to mobility and the factors affecting the unit processes at the interface.

7.6.2. On Atomic Transport

(a) The role of short circuits and quenched-in vacancies on low-temperature reactions should be examined further. A detailed study of the behavior of quenched-in vacancies, their clustering and their disappearance would be valuable.

(b) Another factor which deserves investigation is the effect on the diffusion, if any, of the dislocations created by the transformation strains.

(c) The present situation concerning the kinetic law for the growth of precipitating particles is rather confusing. A method to establish the relevance of Ham's or Zener's solutions is needed. It is felt that a test of stability of shape against fluctuations would provide such a method.

7.6.3. On Complicating Factors in Growth Process

(a) Close scrutiny of the factors which complicate the solid-state growth process seems essential. They should be studied one by one experimentally. Suitable systems where this can be done should be sought.

(b) The role of interfacial energy vs. interface mobility in determining the shape of growing crystals should be studied.

(c) The kinetics of a precipitation process should be investigated before and after the loss of coherency in order to clarify the role of coherency stresses on growth.

7.6.4. On Kinetics of Growth

(a) It is felt that the basic mechanisms of growth discussed in this section must be studied in much more detail before any real advance can be made in this field.

(b) The possible application of several variables separately in order to study the response of a growth process and thus identify the rate-controlling mechanism, should be examined.

(c) The principle which governs the degree to which nonessential steps in a growth process occur should be sought. Related to

this is the important question as to the stage at which local chemical equilibrium can be expected at a phase boundary. This problem needs more attention.

7.6.5. On Lamellar Growth

(a) In studies of lamellar growth, there should be correlated measurements of growth rate, spacing, orientation, and extent of segregation over wide ranges of temperature and composition. This work should be supplemented by thermodynamic and diffusion measurements, if these are lacking.

(b) There is considerable need for systematic microscopic work on the three-dimensional structure of pearlite at different stages of development. High-purity iron-carbon and well-defined ternary alloys should be used as far as possible. A combination of X-ray measurements of orientation and microscopic observation of structure would be especially valuable. The shape of the pearlite-austenite interface should be also studied, using an effective quenching technique and the observations compared with theoretical predictions under different assumptions. In some steels a low-temperature form of pearlite with a ragged contour during growth has been observed. A close examination of this structure and of its relation to ordinary, smooth pearlite may yield fundamental information about the lamellar growth process.

(c) The theoretical treatment of the growth process, assuming grain-boundary diffusion, should be developed further by making the physical model more detailed. A calculation of the shape of the reaction front should be attempted. The influence of alloying elements should also be included in the theoretical treatments of lamellar growth.

(d) Recent work (Hillert, to be published) has indicated that the formation of two phases in a transformation product is not sufficient to produce a lamellar structure, but that an additional factor of "cooperative growth" between the two phases is necessary. It is important to ascertain the conditions leading to such cooperation between the two phases, which permits the start and propagation of a lamellar configuration.

7.6.6. On Bainitic Growth

(a) A thorough investigation of the details of bainitic growth in the pure iron-carbon system should be undertaken. The addition of alloying elements should be considered mainly as a valuable experimental tool and not overemphasized in experiment or theory.

Most experimental work on bainite has previously been carried out on alloyed steels.

(b) Detailed information about the bainitic transformation should be collected, including the variation with temperature and carbon content of the edgewise and sidewise growth rates.

(c) An analysis of the surface relief effects should be carried out, as has been done for martensite.

(d) The conditions under which bainitic structures can form should be examined in a broad sense by investigating a number of eutectoid systems.

(e) A theoretical treatment of the diffusion-controlled growth of bainite is worth attempting. It might supply an explanation for the fact that growth occurs even when the free energy available seems high enough to allow rapid growth without partitioning of carbon.

(f) The change in kinetics at 300–350°C should be examined further and, if possible, related to changes in morphology.

8. GENERAL KINETICS OF SOLID-STATE TRANSFORMATIONS

8.1. Statement of the Problem

One goal in the study of phase transformations is to achieve the prediction of the overall reaction rate. However, the present knowledge concerning the phenomena of nucleation and growth is still too limited to allow the morphology, the rate of nucleation, and the rate of growth to be predicted with any real degree of certainty. The subsequent question of the means of combining this type of information to obtain an overall reaction rate is essentially a problem in spatial geometry, and will not be treated here.

Consequently, the current importance of general kinetics stems from the possibility of turning the objective around and of extracting valuable information concerning nucleation and growth from a knowledge of the overall reaction rate. However, it should be emphasized that, whenever possible, direct studies of the nucleation and growth processes are to be preferred.

The indirect method of investigating nucleation and growth through general kinetics involves two operations: (a) the experimental observation of the progress of the reaction, and (b) the analysis of the kinetic data in terms of various nucleation and growth models. The measurements of the reaction are usually carried out by following the change of some macroscopic physical quantity (for instance, length, electrical resistivity, internal energy, internal friction, or various optical or magnetic properties),

and it is usually assumed that the observed property change is proportional to the volume transformed, an assumption which requires justification. Sometimes, one is able to determine the rate of reaction directly by metallographic methods. In many cases, a satisfactory interpretation of the change of a physical quantity in terms of reaction kinetics is only possible if one considers its relation to all the characteristics of the reaction, such as the particle shape, distribution of the transformation product, and state of strain, as well as the extent of transformation. Despite these uncertainties, the present discussion will deal mainly with the problem of extracting valuable information concerning the details of a transformation from experimental data on the overall reaction rate.

8.2. State of Knowledge

8.2.1. Use of Classical Rate Laws

Attempts to apply the classical rate laws governing homogeneous chemical reactions to nucleation and growth processes are of limited basic value. Even if the experimental data can be fitted to a homogeneous rate equation by properly choosing a certain set of constants for the equation, this practice is of doubtful value, except to facilitate interpolation or extrapolation of the data, and the constants themselves have little or no physical significance.

8.2.2. Evaluation of Rate of Nucleation and Rate of Growth

The overall reaction rate of a nucleation and growth process depends primarily on the rates of nucleation and of growth, with both of these quantities being functions of time in the general case. The volume fraction transformed, f , at time, t , is thus proportional to the integral $\int_0^t \dot{N}(\tau) \cdot \bar{v}(\tau, t-\tau) \cdot d\tau$, where \dot{N} is the nucleation rate, and \bar{v} is the average volume at time t , of the particles nucleated at time τ . It would be highly desirable, indeed, to be able to derive the functions \dot{N} and \bar{v} directly from the overall rate, but this cannot be done without making some assumptions about the nature of the process. Clearly, then, the significance of the resulting information concerning the details of the process will depend on the soundness of the original assumptions. This limits the value of the indirect method of studying nucleation or growth, and is the main reason that direct measurements are preferred, when they can be carried out, even if they are much more laborious.

Unfortunately, only a few simple growth models are well enough understood (sec. 7) to allow the derivation of expressions for \dot{N} and \bar{v} . It is, therefore, quite common to assume that the rates of nucleation and growth, at least for small f , can be represented by simple power functions as \dot{N} proportional to τ^m , and \bar{v} proportional to $(t - \tau)^n$, although more complicated functions have also been tried. The overall reaction rate can then be derived and fairly simple rate laws for small f are obtained if one of the exponents is an integer. In order to evaluate either of them, a certain value must thus be chosen for the other.

To arrive at a value for the growth exponent, it is often assumed that the nucleation exponent is zero, i.e. the rate of nucleation is constant, as predicted by the classical theories for homogeneous nucleation. However, this assumption is of doubtful validity in a real system containing lattice imperfections; in fact, it is often better to assume that all the nuclei are present from the beginning. In other instances, most of the nuclei form during the early stages of transformation, and treating all the nuclei as if they were present from the beginning may then be a good approximation only for the later stages.

When trying to evaluate kinetic data from the later stages of a transformation, another difficulty is encountered due to various disturbances in the growth of the transformation product. This problem is discussed in the following section.

8.2.3. Phenomenon of Impingement

There are several reasons for disturbances in the growth of a reaction product after the very first stage of transformation. Growth may become increasingly difficult as the parent phase becomes depleted in solute, or it may stop completely if the growing particle meets another particle or a grain boundary. It is a difficult mathematical problem to account for such interferences, and the solution has been obtained in only two cases.

(a) The first case to be solved (Kolmogorov, 1938; Johnson and Mehl, 1939) was that concerned with particles growing at a constant rate in all directions until the occurrence of impingement. The nucleation was originally assumed to occur at random nucleation sites and a simple rate law was obtained for a constant nucleation rate. Later, several kinds of preferred nucleation sites were treated (Cahn, 1956), and it was shown (Cahn, 1957b) that the analysis of kinetic data may give quite erroneous results for the

nucleation rate if incorrect assumptions about the rate law and nucleation sites are employed.

(b) The other case for which a satisfactory mathematical solution has been found concerns the diffusion-controlled growth of particles where the dimensions increase as the square root of time before interference occurs (Wert and Zener, 1950; Ham, 1958, 1959). In order to find this solution, one assumes that all the particles are present from the beginning and also that the parent phase is only slightly supersaturated.

When comparing the rate equations obtained with and without the impingement effects taken into account, it is found that the expressions differ only by a factor (called impingement factor) which happens to be identical in the two cases treated so far, $1 - f$, where f is the volume fraction transformed. The common procedure is to introduce the same factor or a similar one in the rate equations for other cases where rigorous mathematical solutions are not available. For instance, the impingement factor has been given different weights according to the interference to be expected from the nature of the reaction, by choosing a certain power of the original factor (Lement and Cohen, 1956).

A fact not always appreciated is that the two rigorous solutions mentioned in (a) and (b) above do not yield the same kind of rate equation, although the impingement factors are identical. In case (a), $\frac{df}{dt} \cdot \frac{1}{1-f}$ is proportional to a certain power of t , and in case (b) to a certain power of f . Hence, there is a problem not only in the choice of exponent for the impingement factor, but also in the choice of rate equation in which to introduce the factor. This is a serious situation because it can be shown that the evaluated growth exponent can be very sensitive to the choice of impingement factor and equation. As the reaction approaches completion, the evaluated exponent becomes less certain. This is clearly demonstrated by the rate equation, which can be derived, especially for the last part of transformation (Ilschner, 1955; Ham, 1958). According to this equation, the growth exponent has no influence at all on the rate of reaction at this stage. Therefore, it becomes quite difficult to extract any information about the growth exponent from data on the later part of the transformation.

8.2.4. Significance of Activation Energy

The preceding discussion has emphasized the difficulties in deriving fundamental information about a reaction by accounting for experimental data in terms of a rate law. Because of these

complexities, a different method is often used which does not aim at determining the details of the reaction, but simply at identifying the rate-controlling mechanism through a study of the temperature dependence of the reaction rate. The procedure does not require any detailed model of the reaction. One simply plots the logarithm of the reaction rate (or logarithm of time of transformation at a certain degree of transformation) vs the inverse of the absolute temperature, and the slope defines an activation energy for the reaction. This is then compared with activation energies for various possible rate-controlling mechanisms, such as diffusion.

However, in order for the evaluated activation energy to be significant with respect to the rate-controlling mechanism, it is necessary that the reaction path be exactly the same for different temperatures, and also that the driving force remain the same. The applicability of this short-cut method may thus be quite limited. Opposing the first requirement is the fact that the number of nuclei may vary sensitively with temperature. Accordingly, it has been suggested (Wert, 1949) that one should allow the reaction to start at a low temperature in order to form a certain number of nuclei, and then study the further growth of these nuclei at various higher temperatures. With regard to the second requirement, the driving force may change very rapidly with temperature, particularly in the neighborhood of the equilibrium temperature. In order to minimize the latter complication, the data should be obtained far away from the equilibrium temperature.

The usual method for testing that the above two requirements are fulfilled is to demand that the plot of \ln rate vs $1/T$ give a straight line over a wide range of temperature, usually extended to low temperatures. However, this test is of doubtful value because it is often difficult to obtain data over a sufficiently wide range of temperature, and furthermore, if a straight line is found, this does not necessarily prove that the requirements are fulfilled.

Sometimes a temperature range can be used even if the requirements are not fulfilled, if the temperature variation of the nucleation rate and driving force can be predicted from a model of the reaction. This method is to be preferred even though a model is needed. On the other hand, it is even better to make the model so detailed that the absolute rate of reaction (not only its variation with temperature) can be tested.

It has been found that the activation energy can vary with the degree of transformation (Gerdien, 1959). Hence, it should be recognized that the reaction to be studied may be of a complicated nature, and that the activation energy at different stages of trans-

formation should be determined, not only at the time for half-completion. A variation of the activation energy with the degree of transformation may indicate that the nature of the growth process is changing. The physical significance of the growth exponent evaluated for such a case is quite unclear.

8.3. Current Ideas

Apart from mathematical difficulties, the real limitation of the indirect method for studying the details of a reaction through the overall reaction rate is that the evaluated quantities are significant only if all the assumptions are correct. The assumptions may never be quite true, since simplifications always have to be used. One way to test the model is to demand that the derived rate equation fit the experimental points over a large portion of the transformation. However, in the present state of knowledge, it seems to be asking too much of a model if it is required to hold throughout the whole reaction.

There has been a tendency, recently, to modify the models to make them fit better. The fact that such attempts have been successful does not always mean that the modified model comes closer to the truth. It may rather be due to the fact that more adjustable constants have been introduced in the rate equation.

A better way to ascertain whether the chosen model is essentially correct is to test it with experimental data of as many kinds as possible. This applies to all aspects of a model that can be checked directly. It has, for instance, been quite common previously to use experimental values of the growth exponent for predictions about the particle shape, assuming rate-control by ordinary diffusion. This procedure breaks down when the experimental growth exponent is less than $\frac{1}{2}$, which is the lowest possible value for ordinary diffusion control. The procedure has also failed in instances where the experimental growth exponent had a reasonable value according to the model, but the actual shape which was directly observed later corresponded to a different value of the growth exponent. It now seems that the shape should be determined by direct microscopic observation, whenever possible, and that this information together with the experimental growth exponent should be used in an attempt to identify the rate-controlling mechanism.

So far, this discussion has related to the extraction of information from experimental data on the rate of reaction. However, there may be changes in several macroscopic properties that may be studied and interpreted in terms of the chosen model. It has

been suggested (Pitsch, 1955; Maurer, 1958) that detailed knowledge about the reaction can be obtained by the simultaneous measurement of two or more such properties. This seems to be a good possibility as long as it is not essential to make a model more detailed simply in order to account for the property observations, without being able to test the details of the model independently.

From the preceding discussion, it is clear that there are many difficulties involved in the basic interpretation of reaction rates. Consequently, very simple cases should be studied, in order to learn more about the nucleation and growth processes from the kinetics. Competing processes, sequential processes, and most polyphase transformations seem to be too complicated to make a closer study of their overall kinetics profitable at the present stage of knowledge. Instead, the elementary steps of such reactions should be studied through more direct measurements.

8.4. Suggestions for Future Work

(a) The nucleation process and the growth process should be studied directly whenever possible, instead of relying on the overall reaction rate for this information.

(b) In cases where only kinetic measurements are available, more attention should be paid to the absolute value of the reaction rate than merely to its variation with temperature or time.

(c) It is possible that the future studies of nucleation and growth, and of the mathematics of impingement, will lead to a better understanding of the overall reaction kinetics, and thereby make it a more useful tool for revealing the details of a reaction than it is at the present time.

9. SUMMARY

In the light of this survey, it is evident that the science of phase transformations has a long way to go before it can be used to foretell the occurrence of solid-state reactions, or to predict their nature, kinetics and morphology. Almost every aspect of solid-state physics and materials science is involved in this subject; and much progress on many fronts is necessary before a genuine understanding of phase transformations can emerge. Nevertheless, it is possible to identify some of the key problems that require critical investigation.

To make a realistic start, a new system of classifying solid-state reactions has been evolved (Table I). This classification is based primarily on growth features, rather than on nucleation features, because the growth process is more amenable to experi-

mental study. Whether or not the growth is attended by a shape change (in contrast to a volume change) becomes an important consideration, inasmuch as it sheds light on the nature of the propagating interface. Moreover, the growth kinetics become significant in revealing whether or not the interface movement is thermally activated. The new classification merits careful scrutiny because it serves to point up many fundamental characteristics of a transformation that require elucidation.

Quantitative data on the chemical driving force of solid-state reactions are quite meager. Notwithstanding the experimental difficulties in obtaining appropriate thermodynamic information at the temperatures involved, this part of the overall problem should be pursued relentlessly, since it bears on the vital question of whether a given transformation can proceed at all. Knowledge concerning the stability of phases, relative to one another in a transformation, is available in comparatively few cases. In addition, there remains the very fundamental question as to whether the thermodynamic properties of bulk phases can be properly applied to the fine-scale states involved in phase transformations.

The nucleation step in solid-state reactions is very sensitive to the exact condition of the parent phase. In this connection, much attention must be directed to such pre-transformation phenomena, as the degree of clustering or short-range order, lattice imperfections, quenched-in vacancies, and the interaction of solute atoms with lattice defects. These subtle factors may have a major influence, not only on the nucleation rate, but on the morphology and reaction kinetics; they also offer an unexploited means of controlling transformations. In fact, it is possible that new transformations may be encountered by modifying the microstate of the parent phase.

At the present time, there seems to be little hope that the interfacial and strain energies at play in solid-state transformations can be measured directly; experimental techniques for quantitatively probing these microscale phenomena are lacking. Instead, there must be reliance on calculations for these quantities, using information about the nature of the interface and the magnitude of the local volume and shape changes. Here, the crystallography of phase transformations is beginning to make a powerful contribution because it enables inferences to be drawn concerning the structure of the interface as well as the atomic movements that participate in the reaction. Martensitic transformations may prove particularly fruitful in this regard inasmuch as the glissile nature of the interface imposes severe restrictions on the possible

mechanisms and on the geometry of the interface. The latter consideration may lead to realistic estimates of the interfacial energy.

In the related problem of strain-energy calculations, account should be taken of coherency effects as well as the volume and shape changes. Unfortunately, the problem is extremely difficult because the condition of anisotropy should not be overlooked. There is now much uncertainty about the origin of the plate-like shape of transformation products; it would certainly be desirable to discriminate between minimization of strain energy, elastic anisotropy, and orientation-dependent growth rates in order to arrive at a unique answer.

The glissile interfaces characteristic of martensitic reactions can respond sensitively to a chemical driving force, because thermal activation of individual atomic movements is not necessary. For other transformations in which thermal activation is required, the interface is said to be nonglissile. The corresponding atomic movements may then be short-range or long-range in nature. The short-range movements consist of atomic jumps across the transformation interface, such as occur in allotropic changes with only one kind of atom, or in order-disorder reactions with more than one kind of atom. The long-range movements consist of atomic transport down an activity (or concentration) gradient over distances of many lattice spacings; this process leads to changes in bulk composition. Considerations of this sort determine the model to be used in analyzing the kinetics of thermally-activated growth. However, quantitative information is almost totally lacking in regard to the concentration gradients at play during a transformation. Conceivably, the electron beam microprobe could turn out to be a valuable tool in this connection.

Due caution should be exercised in the interpretation of overall reaction kinetics. It can prove quite misleading to extract information about nucleation, or growth, or morphology from an assumed kinetic model, without having independent data on at least some of the constituent processes. In the first place, when changes in macroscopic properties are employed to trace the course of a reaction, there is usually imperfect knowledge concerning how such properties quantitatively reflect the extent of transformation. Furthermore, the kinetic interpretation is complicated by impingement effects, short-circuiting diffusion paths, and progressive alterations in the state of the parent phase. Even if the simpler approach is adopted of attempting to analyze the temperature-dependence of the reaction rate instead of the reaction rate itself, the way is paved with uncertainties. The observed

temperature dependence cannot identify the activation energy of the rate-controlling step, unless the transformation path remains the same over the experimental temperature range, and unless the temperature dependence of the driving force (or of the thermodynamic activities involved) is corrected for. There is also the disconcerting possibility that the operative activation energy may actually change as a function of temperature or extent of transformation.

Under these circumstances, more effort should be concentrated directly on the elementary factors that enter into a transformation, such as the state of the parent phase, rates of diffusion, nucleation, and growth, concentration gradients, particle morphology in three dimensions, etc. Then, the overall kinetics (measured directly if possible, or otherwise by more than one bulk property) can be profitably employed to test the kinetic law into which the more basic measurements are fed.

All of the foregoing issues, and more, are pinpointed by the "specific suggestions for future work" listed in each section of this chapter. These suggestions fall into three general categories:

(a) Few, if any, of the fundamental ideas about solid-state phase transformations can be regarded as unchallengeable or fully perfected. Almost every phenomenon operating during the transformation process deserves theoretical study.

(b) The materials selected for experimental investigations should be critically chosen for the purpose. Some obvious examples are high-purity and single-crystal alloys, transparent ionic systems, semiconductors which lend themselves to unique electrical measurement, whiskers or low-defect crystals, etc. At the same time, various types of measurements should be coordinated in a systematic way to insure that reliable and comprehensive information will be brought to bear on the subject.

(c) The applicability of new techniques should be actively pursued (see Part XII on Techniques and Instrumentation). This is probably one of the most challenging approaches to the problem, in view of the present difficulties inherent in observing structure-sensitive microscale phenomena. Transmission electron microscopy, the electron-beam microprobe, X-ray microscopy, microbeam diffraction, and resonance methods are collectively capable of enhancing greatly the real knowledge about phase transformations. Novel ways of introducing lattice imperfections, such as by radiation damage and shock loading, are also of potential importance. Finally, the advent of high-pressure techniques pro-

vides a new dimension for controlling phase transformations and for testing theories.

10. BIBLIOGRAPHY

10.1. Review Articles

- Averbach, B. L., 1956, *Theory of Alloy Phases*, 301. (ASM, Cleveland)
- Barrett, C. S., 1952, *The Structure of Metals*, Second Edition. (McGraw-Hill, New York)
- Beck, P., 1954, *Advances in Physics*, 3, 245.
- Bilby, B. A., and Christian, J. W., 1955, *The Mechanism of Phase Transformations in Metals*, 121. (The Institute of Metals, London)
- Blackman, M., 1955, *Crystal Physics I, Handbuch der Physik*, 325.
- Bowles, J. S., and Barrett, C. S., 1953, *Progress in Metal Physics*, 3, 1. (Pergamon Press, London)
- Bradley, R. S., 1951, *Quart. Revs.*, 5, 315. (London)
- Brooks, H., 1952, *Metal Interfaces*, 20. (ASM, Cleveland)
- Christian, J. W., 1956, *J. Inst. Metals*, 84, 386.
- Cottrell, A. H., 1954, *Relations of Properties to Microstructure*, 131. (ASM, Cleveland)
- Friedel, J., 1954, *Advances in Physics*, 3, 446.
- Guinier, A., 1956, *Trans. AIME*, 206, 673.
- , 1959, *Solid State Physics*, 9, 293.
- Guttman, L., 1956, *Solid State Physics*, 3, 145.
- Hollomon, J. H., and Turnbull, D., 1953, *Progress in Metal Physics*, 4, 33.
- Kaufman, L., and Cohen, M., 1958, *Progress in Metal Physics*, 7, 165.
- Keesom, P., and Pearlman, N., 1955, *Handbuch der Physik VII-I*, 325.
- Lumsden, J., 1952, *Thermodynamics of Alloys*. (Institute of Metals, London)
- Oriani, R. A., 1958, *N.P.L. Symposium on Metallic Solutions*. (London)
- Warren, B. E., and Averbach, B. L., 1953, *Modern Research Techniques in Physical Metallurgy*, 95. (ASM, Cleveland)

10.2. References

- Bainbridge, D. W., Li, C. H., and Edwards, E. H., 1954, *Acta Met.*, 2, 322.
- Basinski, Z. S., and Christian, J. W., 1954a, *Acta Met.*, 2, 101; 1954b, *ibid.*, 2, 148.
- Bilby, B. A., 1953, *Phil. Mag.*, 44, 782.
- Bowles, J. S., and Mackenzie, J. K., 1954a, *Acta Met.*, 2, 129; 1954b, *ibid.*, 2, 138; 1954c, *ibid.*, 2, 224; 1957, *ibid.*, 5, 137.
- Bullough, R., and Bilby, B. A., 1956, *Proc. Phys. Soc.*, B69, 1276.
- Bunshah, R. F., and Mehl, R. F., 1953, *Trans. AIME*, 197, 1251.
- Cahn, J. W., 1956, *Acta Met.*, 4, 449.
- , 1957a, *Trans. AIME*, 209, 140; 1957b, *Acta Met.*, 5, 169; 1959, *ibid.*, 7, 18.
- , and Hilliard, J. E., 1958, *J. Chem. Phys.*, 28, 258; 1959, *ibid.*, 31, 688.
- Cech, R. E., and Turnbull, D., 1956, *Trans. AIME*, 206, 1.
- Christian, J. W., 1958, *Acta Met.*, 6, 377; 1959, *ibid.*, 7, 218.
- Cohen, M., 1958, *Trans. AIME*, 212, 171.
- Crocker, A. G., and Bilby, B. A., 1959, Unpublished work.

- Crussard, C., 1955, *Symposium on the Mechanism of Phase Transformations in Metals*, 309. (Institute of Metals, London)
- Darken, L. S., Fisher, R. M., and Galligan, J. M., 1957, *J. Metals* (Fall Meeting Report 1957) 56.
- DeLauney, J., 1956, *Solid State Physics*, 2, 285.
- DeSorbo, W., and Turnbull, D., 1956, *Acta Met.*, 4, 495.
- Eshelby, J. D., 1957, *Proc. Soc., A*, 241, 376; 1959 *ibid.*, 252, 561.
- Feder, R., Mooney, M., and Nowick, A. S., 1958, *Acta Met.*, 6, 266.
- Flinn, P. A., McManus, G. M., and O'Meara, F. J., 1959, Westinghouse Research Laboratory Report.
- , Averbach, B. L., and Cohen, M., 1953, *Acta Met.*, 1, 664.
- Frank, F. C., 1953, *Acta Met.*, 1, 260.
- Gaunt, P., and Christian, J. W., 1959, *Acta Met.*, 7, 529.
- Gerdien, H. O., 1959, *Archiv f.d. Eisenhw.* (in press).
- Gibbs, J. W., 1948, *Collected Works* (Yale University Press, New Haven)
- Ham, F. S., 1958, *J. Phys. Chem. Solids*, 6, 335.
- , 1959, *J. Appl. Phys.*, 30, 1518.
- Hardy, H. K., 1953, *Acta Met.*, 1, 202.
- Hillert, M., 1960 (to be published).
- Hultgren, A., 1953, *Kungl. Svenska Vetenskapsakademiens Handlingar*, 4th Series, 4, 56.
- Ilshner, B., 1955, *Archiv f.d. Eisenhw.*, 26, 59.
- James, R. W., 1950, *The Optical Principles of the Diffraction of X-rays*. Sir Lawrence Bragg, Editor (G. Bell and Sons, London).
- Jellinghaus, W., Rose, A., and Holetzko, H., 1954, *Archiv f.d. Eisenhw.*, 25, 251.
- Johannson, C. H., 1937, *Arch. f.d. Eisenhw.*, 11, 241.
- Johnson, W. A., and Mehl, R. F., 1939, *Trans. AIME*, 135, 416.
- Kaufman, L., 1959a, *Acta Met.*, 6, 575.
- Kaufman, L., 1959b, *Bull. Am. Phys. Soc.*, 114, 181 (Technical Report No. 2, Nour 2600 (00) Office of Naval Research).
- Kelly, P., and Nutting, J., 1959, Unpublished work.
- Kimura, H., Maddin, R., and Kuhlmann-Wilsdorf, D., 1959, *Acta Met.*, 7, 145.
- Kirkaldy, J. S. (to be published).
- Kolmogorov, 1938, Published by Mirkin, L. L. *Sbornik Trudov Moskovskogo Instituta Stall*, No. 10.
- Lement, B. S., and Cohen, M., 1956, *Acta Met.*, 4, 469.
- Lieberman, D. S., 1958, *Acta Met.*, 6, 680.
- Massalski, T. B., 1958, *Acta Met.*, 6, 243.
- Maurer, R. D., 1958, *J. Appl. Phys.*, 29, 1.
- Munster, A., and Sagel, K., 1958, *N.P.L. Symposium on Metallic Solutions* (London).
- Nabarro, F. R. N., 1940, *Proc. Phys. Soc.*, 52, 90.
- Nicholson, R. B., Thomas, G., and Nutting, J., 1959, *J. Inst. Metals*, 37, 429.
- Normann and Scheil, 1958, *Progress in Metal Physics*, 7, Private Communication.
- Oriani, R. A., 1957, *J. Phys. Chem. Solids*, 2, 327.
- Otte, H. M., and Read, T. A., 1957, *Trans. AIME*, 209, 412.
- Pashley, D. W., and Presland, A. E. B., 1959, *J. Inst. Metals*, 37, 419.
- Pitsch, W., 1955, *Acta Met.*, 3, 542.
- Powers, R. W., and Doyle, M. V., 1959, *J. Appl. Phys.*, 30, 514.
- Rayne, J., 1957, *Phys. Rev.*, 108, 22.
- Seeger, A., 1953, *z. Metallk.*, 44, 247.

- Seeger, A., 1956, *z. Metallk.*, **47**, 653.
Speich and Cohen, 1960 (to be published).
Suzuki, T., 1954, *Sci. Rep. Res. Inst. Tohoku Univ.*, **8**, 309.
Takeuchi, S., and Honma, T., 1957, *Sci. Rep. Res. Inst. Tohoku Univ.*, **9**, 492 and 508.
Turnbull, D., 1956, *Solid State Physics*, **3**, 226.
———, Rosenbaum, H. S., and Treafis, H. N., "Kinetics of Clustering in Some Aluminum Alloys" (to be published).
Wechsler, M. S., Lieberman, D. S., and Read, T. A., 1953, *Trans. AIME*, **197**, 1503.
Wei, C. T., Cheng, C. H., and Beck, P. A., 1959, *Phys. Rev. Letters*, **3**, K539.
Weiss, R. J., and Tauer, K. J., 1956, *Phys. Rev.*, **102**, 1495.
———, and ———, 1958a, *J. Phys. Chem. Solids*, **7**, 249.
———, and ———, 1958b, *J. Phys. Chem. Solids*, **4**, 135.
Wert, G., 1949, *J. Appl. Phys.*, **20**, 943.
———, and Zener, C., 1950, *J. Appl. Phys.*, **21**, 5.
White, J. L., Orr, R. L., and Hultgren, R., 1957, *Acta Met.*, **5**, 747.
Zener, C., 1946, *Trans. AIME*, **167**, 550.

Part VII

GROWTH, STRUCTURE, AND MORPHOLOGY OF CRYSTALS

Chairman

Dr. DAVID TURNBULL
General Electric Co.
Schenectady, N. Y.

Members

Dr. LEO MANDELKERN
National Bureau of Standards
Washington, D. C.

Prof. R. V. COLEMAN
University of Virginia
Charlottesville, Va.

Dr. J. S. KASPER
General Electric Co.
Schenectady, N. Y.

Dr. MORRIS TANENBAUM
Bell Telephone Laboratories
Murray Hill, N. J.

Dr. W. W. WEBB
Cornell University
Ithaca, N. Y.

Dr. J. W. MENTER
Tube Investment Research
Laboratories
Hinxton Hall, England

CONTENTS

	Page
1. INTRODUCTION AND ACKNOWLEDGMENTS.....	385
2. HISTORICAL SURVEY	387
2.1. Early Period	387
2.2. Middle Period	388
2.3. Recent Period	389
3. CRYSTAL CHEMISTRY	390
3.1. Definition and Scope of Crystal Chemistry.....	390
3.2. Classification of Structures According to Type of Bonding.....	391
3.3. Special Problems Arising in the Application of Crystal Chemical Principles	392
3.3.1. The Question of Atom Size.....	392
3.3.2. Failures of Radius Ratio Criteria.....	392
3.3.3. Uncertainty of Bonding Type.....	392
3.3.4. Structure Distortions	393

	Page
3.3.5. Non-Crystallographic Coordinations	393
3.3.6. The Crystal Chemistry of Metals	393
3.3.7. Non-Crystalline Substances	394
3.4. Recent Advances in Crystal Chemistry	394
3.4.1. Advances Concerning Geometry	394
3.4.2. Advances in Knowledge of Binding Forces	395
3.4.3. Novel Structure Types	396
3.4.4. Protein Structures	397
4. THEORY OF GROWTH AND MORPHOLOGY	397
4.1. Structure at the Crystal Surface	397
4.2. Growth into Dilute Fluids	399
4.2.1. Dislocation Mechanism	399
4.2.2. Etch Pits	399
4.2.3. Electrodeposition	400
4.2.4. Impurity Effects	400
4.2.5. Kinematic Theory of Crystal Growth	400
4.2.6. Crystal Dissolution	401
4.3. Growth into Concentrated Fluids	401
4.4. Boundary Migration in Crystalline Solids	402
4.4.1. Nuclei for Recrystallization, Primary and Secondary	402
4.4.2. Crystal Boundary Migration	403
4.5. Crystal Morphology	405
4.5.1. Dendrites	405
4.5.2. Spherulites	406
4.5.3. Whiskers	406
5. CONTROL OF PURITY AND STRUCTURE	407
5.1. Crystal Growing Techniques	407
5.1.1. General Experimental Problems	408
5.1.2. Crystal Growing Methods—Requirements and Limitations	411
5.2. The Control and Determination of Impurity Concentrations in Solids	415
5.2.1. Impurity Determination	416
5.2.2. The Equilibrium Diagram	417
5.2.3. Thermodynamic Properties	418
5.2.4. Kinetics	419
5.2.5. Interactions of Impurities with other Lattice Defects	419
5.3. Perfection Control	420
5.3.1. Determination of Dislocation Structure	421
5.3.2. Control of Dislocation Structure	424
5.4. Synthesis of Substances into Different Crystal Structures	428
6. CRYSTALLIZATION OF POLYMERS	430
6.1. Introduction	431
6.2. Structure and Morphology	432
6.3. Mechanism of Liquid-to-Crystal Transformation	433
7. FORMATION AND STRUCTURE OF THIN FILMS	435
7.1. Production of Reproducible Substrates for Film Growth	435

	Page
7.2. Characterization of Defects in Thin Films by Electron Microscopy	437
7.3. Oriented Overgrowth and Growth of Single Crystal Films	439
7.4. Microstructure of Single Crystal Films	441
7.5. Thin Film Properties	442
8. SUMMARY OF CONCLUSIONS AND CURRENT IMPORTANT PROBLEM AREAS	444
8.1. Crystal Chemistry	444
8.2. Theory of Growth and Morphology	445
8.3. Control of Purity and Structure	446
8.3.1. Crystal Growing Techniques	446
8.3.2. The Control and Determination of Impurity Concentrations in Solids	446
8.3.3. Perfection Control	447
8.3.4. Synthesis of Substances into Different Crystal Structures	448
8.4. Crystallization of Polymers	448
8.5. Formation and Structure of Thin Films	449
9. REFERENCES	450
9.1. Reviews and General Articles	450
9.2. Original Articles	452

GROWTH, STRUCTURE, AND MORPHOLOGY OF CRYSTALS

1. INTRODUCTION AND ACKNOWLEDGMENTS

The scope of this study is indicated fairly well by the title. In our definition of structure, however, we have included the defect configuration as well as the structure of the unit cell. In fact, we have given considerably more attention to the subject of defect structures than to other structural problems.

After presenting a brief historical survey of the field (sec. 2), we begin the discussion of the current state of knowledge by considering the factors which determine the configuration of the unit cell (sec. 3). Although this same problem is treated at a more fundamental level elsewhere in this volume (see Part II), there exists a body of empirical generalizations in the field of crystal chemistry which has proved extraordinarily useful in correlating crystal structure data and in predicting the stability of substances in various structures; it is this empirical body of knowledge to which we have given most attention in this report.

The next section of the report (sec. 4) is concerned with the theory of crystal growth and morphology. Since crystal growth is essentially an interfacial phenomenon, the basic problem is to develop a satisfactory theory of interfaces. Thus, if we could characterize precisely the structure of the interface, and if we

understood in detail the various processes and interactions which occur at interfaces, in both high-purity and impure materials, we should be able to explain the mechanism of crystal growth, the morphology of crystals, and the defect structures resulting from growth. Indeed, some of the major advances in the theory of crystal growth have resulted from the application of our understanding of interfacial phenomena. Nevertheless, much of our knowledge concerning crystal growth is still embodied in empirical generalizations not yet explained by the theory of interfaces. In the present discussion, we have considered the state of knowledge at this empirical level as well as at the more fundamental level of atomic interactions at interfaces.

With a proper understanding of crystal growth, we should also be able to evaluate the possibilities for synthesizing a given substance in various prescribed structures. According to our definition of structure, the synthesis of diamonds, the growth of single crystals, the preparation of ultra-pure and dislocation-free crystals, are all regarded as examples of structure synthesis. Two of the main parts of our report, sections 5 and 7, are concerned with the present state of understanding of structure synthesis.

Finally, we have included a discussion of the crystallization of polymers, which appears to be a complicated phenomenon; this is indicated by the fact that it is difficult to grow isolated polymer single crystals of appreciable size. However, we believe that in the future the field of crystalline polymers will prove to be an exceptionally fruitful area for solid state investigations. The results of such studies may also have an important impact on our understanding of biological phenomena. We are therefore devoting one main section (sec. 6) of our report to the subject of polymer crystallization.

Recently the status of the subject, with the exception of the crystal chemical aspect, was discussed quite completely at an international conference on crystal growth held in Cooperstown, N.Y. We have made extensive use of the published proceedings ["Growth and Perfection of Crystals", hereafter referred to as GPC, 1958R] of this conference during the course of our study.

In preparing this report we have also consulted with many of our colleagues. We are especially indebted to C. D. Thurmond who contributed the section on the control of impurity content in crystals and to L. Himmel for valuable advice on many matters.

In addition, we have benefited greatly from the advice and criticism of:

G. A. Bassett	L. R. Johnson	W. L. Roth
M. Blackman	G. T. Kohman	J. W. Rutter
S. S. Brenner	R. A. Laudise	G. W. Sears
E. Buehler	J. M. Lommel	A. Smakula
N. Cabrera	J. W. Mitchell	R. G. Treuting
P. H. Egli	E. W. Müller	A. Tweet
W. C. Dash	J. W. Nielsen	J. Washburn
C. G. Dunn	D. W. Pashley	J. H. Wernick
W. E. Ellis	F. P. Price	W. J. Zimmerman, III
S. Geller	H. S. Rosenbaum	

2. HISTORICAL SURVEY ¹

It is convenient for the purposes of our discussion to distinguish three periods of development of the subject. There is the first or early period in which crystal scientists were preoccupied with the concept of the ideal crystal. We suppose that this period ended about the time of the first successful application of X-ray diffraction methods to the analysis of crystal structures. Then follows the middle period in which perhaps the most conspicuous development was the appreciation and elucidation of the role of imperfections in crystal processes. The end of this period might logically be marked by Frank's invention of the screw dislocation mechanism of crystal growth. It is not yet possible to delineate the third and most recent period as definitively as the previous two, but perhaps its most significant feature has been the achievement of important structure syntheses.

2.1. Early Period

During this period, a highly idealized concept of crystals was prevalent, and the importance of crystal imperfections generally was not appreciated. Indeed, many of the structure-insensitive properties of crystals were interpreted satisfactorily using the perfect crystal as a model. Quite naturally, the main preoccupation of crystal scientists at the time was with unit cell configurations and with macrostructural features such as the external form of crystals.

The contributions of Gibbs to the theory of crystal interfaces are among the most important in this period. These contributions

¹ An excellent perspective on the history of important aspects of the field was given by F. C. Frank in his introductory lecture at the Cooperstown Conference (GPC, 1958a).

stimulated important developments in the theory of the equilibrium growth forms of crystals and also laid the foundations for later important advances in the theory of crystal growth. Gibbs clearly enunciated the concept of solid surface tension and stated that the equilibrium shape would be that for which the surface free energy is a minimum. He also recognized that a large energy fluctuation, associated with the formation of a two-dimensional nucleus on the crystal surface, would be required for the growth of perfect crystals under small driving forces. Little was done to elaborate on this idea for a long time, probably because the necessary framework for handling kinetic problems of this type was lacking.

Following Gibbs, Wulff and Curie showed how the equilibrium shape of the crystal could be found from the variation of surface tension with the crystallographic orientations of the bounding planes. Curie advanced the opinion, which became widely accepted, that actual growth forms are dictated by equilibrium requirements. However, the driving free energy for a shape change in macrocrystals is only a minute fraction of the thermal energy. Therefore, the growth form should in general be determined by kinetic rather than thermodynamic conditions. Later, it was demonstrated that a crystal becomes bounded during growth by faces which develop perpendicular to the direction of *slowest growth*. According to the earlier generalization of Bravais, to which many exceptions have been found, the bounding faces of actual crystals are those in which the molecular packing is densest. Generally, but not always, the growth rate is therefore a minimum in directions normal to the most closely packed faces.

Much of the experimental work in this period was directed toward the control of macrostructure. Basic procedures were developed for growing single crystals and eliminating macrostructural flaws.

2.2. Middle Period

As already implied, the major development which characterizes or distinguishes this period was the recognition that the difference between perfection and near perfection is the decisive variable in many crystal processes (e.g., diffusion and plastic deformation). We have stated the point in this manner in order to emphasize that certain phenomena—the structure-sensitive properties—were explained satisfactorily by invoking only minute departures from perfection, for example, one atom disarranged in 10^6 .

The first conclusive evidence for the existence of crystal im-

perfections was probably derived from the early X-ray investigations. Then, in the early 1920's came the pioneering work of Frenkel, Schottky and Wagner on the role of point defects, both structural and chemical, in solid state processes. This work probably provided the initial stimulus to efforts that led to the synthesis of ultra-pure crystals. Later, by about 1930, Taylor, Orowan and others showed that the qualitative features of the mechanical behavior of crystals could be explained by the motion and interactions of dislocations. It was realized that, because of their high energy, dislocations should not exist in crystals of ordinary size at *equilibrium*. Therefore, it was supposed that dislocations usually originated from accidents of growth during the crystal growing process. A comprehensive theory of dislocation structures, which introduced the concept of the screw dislocation, was formulated by J. M. Burgers in 1939.

An important aspect of the middle period was the development of the body of empirical generalizations that constitute the present domain of crystal chemistry. These generalizations proved extraordinarily successful in correlating the structures of ionic and covalent crystals. Zachariasen also applied them, with outstanding success, to account for glass formation in ceramics and salts.

Another major achievement during this period was the development of reaction rate theory. This provided the basis upon which Volmer, Becker, Frenkel and others were able to develop kinetic theories for nucleation and growth.

Not long afterward, it was recognized that a large driving force would be required for the growth of crystals bounded by densely packed planes. In view of the observations that crystals actually grow perceptibly under very small driving forces, it was supposed by some (e.g., Volmer, G. P. Thomson) that the surface must somehow be disordered at equilibrium. However, in 1947, Burton and Cabrera showed conclusively that, at equilibrium, a densely packed crystal surface in contact with a dilute fluid at low temperature should be virtually perfect. As is well known, Frank resolved the dilemma created by this result when he proposed the screw dislocation mechanism of crystal growth.

The Frank theory and the semiquantitative success of the dislocation theory in other areas stimulated efforts to detect dislocations in crystals directly. These efforts were successful and the results, among other things, vindicated the Frank theory.

2.3. Recent Period

As stated above, the recent period seems best characterized by

notable successes in structure syntheses. We note first the development by Pfann of zone-refining techniques for the preparation of ultra-pure materials. These techniques now have been applied successfully to a wide variety of substances—metals, inorganic salts, organic compounds—and to the growth of ultra-pure crystals.

The outstanding successes of the dislocation theory stimulated interest in the possibility that dislocation-free crystals might be synthesized. In all probability, it was this interest that led to the recognition that crystal bodies which are very small in one or more dimensions (i.e., droplets, filaments or "whiskers", and thin films) sometimes behave in a manner similar to that predicted for dislocation-free crystals. It was expected that such small bodies might either contain dislocations in unusual configurations or none at all.

The goal of synthesizing bulk, dislocation-free crystals was, apparently, first achieved by Dash, who was able to grow, from the melt, perfect crystals of elemental silicon. By using a recrystallization technique, Mitchell has since succeeded in growing bulk silver chloride crystals which are free of dislocations.

Also in the recent period, techniques for achieving and maintaining extremely high pressures at high temperatures were developed by Bundy, Hall, Strong, and Wentorf. Application of these techniques led to the synthesis of diamond and of boron nitride in the diamond modification.

In the polymer field, procedures were developed for growing isolated polymer single crystals about 0.1 micron thick and several microns wide. This ranks as an important achievement, considering that crystallization of polymers usually results in an aggregate of many crystallites none of which is larger than 0.1 micron in any dimension.

3. CRYSTAL CHEMISTRY

3.1. Definition and Scope of Crystal Chemistry

Crystal chemistry deals with the interrelationships between the crystal structures (or atomic arrangements) and the physical and chemical properties of substances. It is concerned with the geometrical aspects of crystallography, with the chemical nature of the constituent units of crystals (atoms, molecules or ions), and with the types of interatomic forces which exist in crystals. The basic postulates of crystal chemistry which are used in establishing structure correlations may be stated as follows:

- (1) A size (atomic radius) can be assigned to a given atom (or ion or molecule).
- (2) The coordination about a given atom (ion or molecule) is of prime importance and is determined by:
 - (a) Geometry and symmetry. The sizes of the units impose a restriction as to the number of possible contacts. A restriction on the nature of the configuration about a given atom is imposed by requirements regarding space-filling and the generation of repeating structural elements.
 - (b) Type of binding forces.

In the practice of crystal chemistry, physical and chemical knowledge is often used to classify and interpret the different crystal structure types; conversely, the detailed findings of a structural investigation often clarify uncertain points in the chemistry and physics of solids and contribute to a better understanding of the nature of the binding forces.

Thus, the scope of crystal chemistry is extremely broad, encompassing all types of atomic arrangements and wide areas of chemistry and physics. On the other hand, it is also empirical in character and its descriptions frequently are of a qualitative rather than a quantitative nature. Nonetheless, marked success has been achieved in correlating structure with physical and chemical properties. More significantly, there has been a fair measure of success in the prediction of structures, not only for known compounds prior to an investigation of their atomic arrangement, but also for postulated compounds before their synthesis has been accomplished. Accordingly, crystal chemical principles offer perhaps the best guides for the synthesis of new substances with specific desired properties.

3.2. Conventional Classification of Structures According to Type of Bonding

The classification of structures into four major categories according to bond type is well known, and detailed treatments have been published by Pauling (1960R), Evans (1939R), Wells (1950R) and Kittel (1956R), among others. This classification, restated here merely for completeness, is as follows:

- (1) Ionic Structures
- (2) Covalent or Homopolar Structures
- (3) Metallic Structures
- (4) Molecular Structures

(5) Special Structure Types

- a. Hydrogen-bonded structures
- b. Interstitial compounds

3.3. Special Problems Arising in the Application of Crystal Chemical Principles

Despite the satisfactory state of development of crystal chemistry, there remain a number of deficiencies and problems; these are briefly noted in order to indicate some areas worthy of further attention.

3.3.1. The Question of Atom Size

Atomic or ionic radii are empirical quantities which have been used to account for the interatomic distances of a vast number of substances on the simple assumption of additivity. It is inevitable, therefore, that discrepancies should arise in specific instances. Nonetheless, for ionic substances, no real need seems to have arisen for revising the commonly used tables of radii prepared by Goldschmidt, Pauling, and Zachariasen. A problem still to be solved, however, is to elucidate the true meaning of atomic size in metal structures and the extent to which standard metallic radii can be employed in dealing with complex alloy structures.

3.3.2. Failures of Radius Ratio Criteria

In ionic substances, one of the main factors which determine the coordination and the type of structure is the ratio of the cation to anion radius. While this has proved to be a good criterion in general, it has met with some outstanding failures, e.g., when used to predict whether AB compounds should have the sodium chloride or the cesium chloride structure.

3.3.3. Uncertainty of Bonding Type

Some examples are:

- (a) Substances with the NiAs structure.
- (b) Substances with the zinc blende and wurtzite structures.
- (c) Combinations of metallic elements that are widely separated in the periodic table.
- (d) SiO_2 and silicate structures.

Electronegativity criteria have been used for estimating the amount of ionic character of a bond. This approach presents various difficulties, and it would be desirable if current schemes

for assessing the character of bonds could be improved and made more quantitative.

3.3.4. Structure Distortions

Classical crystal chemistry has not been capable of explaining deviations from a highly symmetric structure that appears to be the rule in any class of related substances. Some deviations of this kind, particularly for transition metal ions, have now been explained by crystal field theory and the Jahn-Teller effect (Dunitz and Orgel, 1957). However, many unexplained phenomena still remain, e.g., perovskite-like structures of non-cubic symmetry.

3.3.5. Non-Crystallographic Coordinations

Coordination geometries such as those of an icosahedron have not been considered previously because of their non-crystallographic nature. Such geometries are now known to occur in a variety of structures, particularly those encountered in complex alloys and inert-gas hydrates.

3.3.6. The Crystal Chemistry of Metals

Aside from the general expectation that metal structures should have high coordination numbers, little guidance of a crystal chemical nature is available in dealing with metal structures. The seemingly promising approach of Brillouin-zone analysis, as developed by Jones (1934) in justification of the Hume-Rothery rules, has not proved satisfactory, except possibly when applied to simple structures such as the brasses. Much remains to be done in clarifying the role of electronic factors and in understanding the geometries of complex structures.

There are differences of opinion even as to the qualitative characterization of the binding forces in metals and alloys. The treatment of Pauling (1955R) contrasts sharply with the band treatments preferred by most physicists. Pauling's approach allows for interpretation of known structures but not for the prediction of new structures.

A problem that invites experimental decision but which has not yet been successfully solved is that of electron transfer in alloys. Although agreement often exists that electron transfer occurs in specific cases, conflicting opinions frequently arise concerning the direction of the transfer, as well as its magnitude (Pauling, 1955R; Laves, 1955R).

All in all, then, metal structures constitute a major area where more effort is needed to develop better crystal chemical principles.

3.3.7. Non-Crystalline Substances

It is within the province of crystal chemistry to consider the liquid and glassy states and other states lacking three-dimensional periodicity, but the understanding, from a structural standpoint, of liquids and glasses is very limited. More structural investigations and new concepts seem to be called for, and this appears to be another major area for crystal chemical advances.

3.4. Recent Advances in Crystal Chemistry

Only brief recognition can be given to some of the recent developments which appear noteworthy because of their more general applicability to crystal chemistry. There are, of course, many illustrations of detailed and refined advances that have been made in the crystal chemistry of an isomorphous or related series of compounds. One such example is the large class of materials of the zinc blende structure that are of interest because of their semiconducting properties. Another is the class of ferrimagnetic garnets containing rare-earth ions. In both cases, previous crystal chemical principles have been useful in the initial stages, but with more detailed study, extensions of the simple principles, specific to the materials of interest, have gradually emerged.

3.4.1. Advances Concerning Geometry

While the geometrical aspects of crystal chemistry have usually been regarded as well-established and understood, there have been new and significant developments, sufficient to suggest, in fact, that the geometrical problems are not all worked out and that geometrical factors may be more important than generally assumed.

An extensive treatment of the spatial geometry of networks has been given by Wells (1958), in which the topological features are emphasized; several new types of networks (with low coordinations) are described, and the relationships of various networks to different structure types are explored. An assessment of the full significance of this work has not been made, but it is evident, at least, that new geometrical features pertinent to crystal chemistry have been disclosed.

Another contribution of a geometrical character concerns complex structures with high coordination numbers such as are observed in various metals and alloys (Frank and Kasper, 1958). A large class of such structures (including σ -phases, Laves phases, β -tungsten structures) have one feature in common, namely, that

only tetrahedral interstices are present; this suggests efficient packings of somewhat unequally sized (but not too greatly different) spheres or slightly deformed spheres. New packing principles for such structures are deduced allowing for the incorporation of non-crystallographic coordinations, such as that of the icosahedron. While metallic structures have been the main concern of this treatment, it has been shown that the same principles can be applied to the inert-gas hydrate structures, which are complex, open structures. The inert-gas hydrates contain non-crystallographic polyhedra which are the mathematical duals of those found in the alloy structures, and hence a strict relationship can be established between these seemingly different kinds of structures.

Another example of a somewhat general nature, and which pertains to geometry, is the layer analysis of metal oxide structures by Iida (1958) (see also Loeb, 1958). Many of the features discussed have undoubtedly been recognized by others, but the treatment given by Iida is more unified than any developed heretofore, and a simple and useful symbolism is presented. The ways of partitioning a hexagonal lattice are worked out, and the point is made that the stoichiometric formula is achieved with the minimum number of layers. Since many oxides do consist essentially of close-packed arrays of oxygen ions, the discussion covers a large number of structures. It may be extended to other salts where close packing of anions occurs. There are, of course, many structures where the anions are not close packed, and these are outside the scope of Iida's analysis.

3.4.2. Advances in Knowledge of Binding Forces

A most outstanding contribution to our knowledge of the binding forces in crystals has been provided by the development and application of crystal field theory. While the basic principles of crystal field theory were formulated a long time ago (e.g., the work of Bethe in 1929), it is only in the past 5 or 6 years that proper appreciation of the subject has arisen. Application of the theory in a crystal chemical sense has been largely due to Dunitz and Orgel (1957A, 1957B) and McClure (1957). Two of the points explained are: 1) the distortions from regular octahedral symmetry for compounds containing Cu^{2+} , Mn^{3+} and Cr^{2+} , where the Jahn-Teller effect is especially prominent; 2) the distributions of different cations between the octahedral and tetrahedral sites in spinel structures.

Previously, neither of these phenomena could be accounted for satisfactorily, although an interpretation for the distortions about

Cu^{2+} had been given by Pauling and an attempt to explain the other distortions was made by Goodenough and Loeb (1955). These interpretations were based on postulated hybridized types of bonding. The crystal field explanations are now more widely preferred. There is a much wider scope to the application of the theory than indicated by the examples given above (Moffitt and Ballhausen, 1956R). In particular, it has been applied extensively to the numerous inorganic complexes of the transition elements.

A new type of bonding has been postulated for the boron hydrides by Eberhard-Crawford, and Lipscomb (Lipscomb, 1959), namely, a three-center bond, wherein a triangular arrangement of atoms is stabilized by an electron pair. While this type of bond seemed to be special to boron hydrides, where it occurred only intramolecularly, it has now been found in a new crystalline modification of boron, indicating that it is necessary to consider this bond as contributing to the crystal forces.

A special type of bonding ("half-bonds") has been postulated (Rundle, 1957) for interstitial compounds of metals with C or N, in order to account for the change in the arrangement of the metal atoms which accompanies the formation of carbides or nitrides. The same type of bonding has been used to explain the stability of various electron-deficient molecules and compounds.

3.4.3. Novel Structure Types

Several examples may be cited of structures that appear sufficiently novel that they may present new problems, either of a geometrical nature, or in regard to the nature of the binding forces involved.

The clathrate crystals wherein cage frameworks are formed and neutral molecules are trapped have now been studied rather extensively, but there is much to be learned about the forces involved and the architectural principles governing their stability. The inert-gas hydrates are especially interesting because of the noncrystallographic nature of the cages formed by the water molecules.

The structures of boron and borides remain puzzling. For some unknown reason, boron has a tendency to assume icosahedral configurations (Hoard et al., 1958; Decker and Kasper, 1959), such as have been found in both crystallographic modifications analyzed to date, as well as in boron carbide. Again, problems exist both with regard to the binding and to the geometry. The bonding theory needs to be developed further, and it is not known how to construct possible three-dimensional networks with the icosahedra.

Much attention has been given to ferrocene and related sandwich compounds because of the implication that new kinds of bonding may exist and because of the unusual geometric arrangement. The problem of binding is still not resolved.

The structure of white phosphorus consists of P_4 tetrahedra distributed at the sites of the α -Mn structure. Since the α -Mn structure itself is not understood, the structure of white phosphorus is all the more perplexing.

Magnetic structures, i.e., the arrangements of the magnetic spins of atoms or ions with unpaired electrons, may be mentioned in this connection, particularly as they effect structure distortions. Thus, MnO is cubic at room temperature but becomes slightly rhombohedral as a consequence of the transformation from a paramagnetic state to an antiferromagnetic state at $\sim 120^\circ\text{K}$.

3.4.4. Protein Structures

The postulate of a nonintegral helix for polypeptides and fibrous proteins by Pauling, Corey and Branson (1951) must be singled out as a most noteworthy advance which has accelerated research in this area. Much has already been learned of the stereochemical principles, and other significant developments can be expected in the near future. The advances in knowledge of protein structures may contribute greatly to our understanding of the structures of other materials such as polymers and hydrogen-bonded molecular crystals.

4. THEORY OF GROWTH AND MORPHOLOGY

Clearly, if we understood in detail the atomistic mechanism of crystal growth, we should be able to predict the crystal morphology. At present, however, we lack a detailed understanding of the growth mechanism, and, in fact, observations on morphology often have stimulated developments in the theory of crystal growth.

As we stressed in the introduction, the mechanism of crystal growth depends largely upon the atomic structure of the interface between the crystal and the medium in which it grows. We will now attempt to summarize our knowledge of the structure of crystal interfaces.

4.1. Structure at the Crystal Surface

As yet, a rigorous and complete theory of the structure (molecular and electronic) and energy of crystal interfaces has not been developed. However, through the use of thermodynamics, certain useful conclusions have been drawn concerning the phe-

nomenology, and additional conclusions concerning the atomistic structure have emerged mainly through application of the nearest-neighbor model of intermolecular cohesion.

We shall consider first the structure of crystal surfaces in contact with dilute (relative to the crystal components) fluid media. By application of nearest-neighbor models, the following main conclusions have emerged concerning the *equilibrium* structure of crystal surfaces at temperatures far below the melting temperature:

- (1) Crystals should be bounded by the planes of highest molecular or atomic density.
- (2) The surfaces of relatively high packing density should be smooth and virtually perfect on an atomic scale. Thus, the transition from crystal to fluid should occur very abruptly, i.e., within one or two atomic distances. The justification for this conclusion is developed in detail by Burton, Cabrera and Frank (1951).
- (3) There should be no sensible density of line imperfections emergent in the surface.

The manner in which these conclusions should be modified for temperatures near the melting point is still not clear.

If, as a result of the growth methods or processing variables, a crystal actually comes to be bounded by nonequilibrium faces, it will tend to change its shape so as to decrease its surface energy. During this process, a stepped or hill-and-valley surface structure may develop in the manner described by Herring (1951). This hill-and-valley structure, which may have a period no more than a few atoms wide, develops so as to expose ledges and faces corresponding to cusps in the Wulff polar diagram of surface free energy vs crystal orientation.

The problem of the structure of the interface between the crystal and its melt, or between a crystal and a concentrated solution of its component elements, is still largely unresolved. For close-packed surfaces in contact with the concentrated fluid, it is useful to distinguish between two limiting possibilities: (1) "singular" interfaces which are characterized by a cusp in the Wulff diagram, in which case the crystal surface would be virtually smooth on the atomic scale, corresponding to an abrupt transition from the crystal to the fluid; (2) "nonsingular" interfaces characterized by shallow minima, but not a cusp, in the Wulff diagram, for which the interface would be somewhat diffuse and the crystal surface rough at equilibrium.

Recently the problem of the diffuseness of interfaces has been

treated thermodynamically by Cahn and Hilliard (1958); (see also Hart (1959) for a general formulation). From the theory developed by Cahn and Hilliard, both the equilibrium thickness of the transition region and the surface free energy can be evaluated knowing the free energy functions for all states intermediate between the two homogeneous phases on either side of the interface. The theory has been applied successfully to the fluid-fluid case and to the crystal-crystal case where the structure is the same on both sides of the interface. Here it is found that the interface becomes progressively more diffuse with increasing temperature. At present, the application of the theory to the crystal-fluid problem would be difficult.

4.2. Growth into Dilute Fluids

4.2.1. Dislocation Mechanism

Gibbs pointed out that the advancement of a singular crystal face of a perfect crystal could occur only by a succession of two-dimensional nucleation events. However, the full implication of this statement was not recognized until many decades later when the necessary theoretical framework for dealing with the kinetics of growth was developed. Burton, Cabrera and Frank then gave the clearest formulation of the problem and showed that singular faces in perfect crystals should not grow at sensible rates unless there are comparatively large departures from equilibrium. As is well known, Frank (1949) was able to reconcile this with the vast amount of seemingly contradictory experience by his screw dislocation mechanism of crystal growth. The dislocations required by this mechanism should not, as we have seen, exist in the crystal at equilibrium. They must originate from various accidents of growth, the nature of which has been recognized in some situations but not in others.

The Frank theory has not only stimulated many experiments but it has also had a tremendous impact on further theoretical developments in the field of crystal growth during the last 10 years. It now appears that the theory is quite well vindicated for the conditions under which it was meant to apply. We will now take up some of the recent developments in the theory.

4.2.2. Etch Pits

Cabrera and Levine (1956) and Cabrera (1957) have extended the theory to take into account the effects of strain energy on the configuration around the dislocation. They show that during

evaporation a hollow-core dislocation will develop into a pit at a certain critical undersaturation. Experimental confirmation of these ideas is still lacking.

4.2.3. Electrodeposition

The kinetics of metal electrodeposition have been interpreted on the basis of the screw dislocation theory, particularly by Vermilyea (1956). Certain qualitative features of the overpotential and of the kinetics of discharge of metal ions on metals were shown to be consistent with the theory.

4.2.4. Impurity Effects

Small traces of certain impurities can have remarkable effects on both the growth rate and morphology of crystals (Buckley, 1951R). These impurities may affect growth by interactions with the crystal surface or by being incorporated into the crystal. In some situations, small traces of impurities decrease the growth rate by several orders of magnitude. With the help of our new understanding of crystal imperfections and related phenomena, considerable progress has been made recently in unraveling some of the complex phenomenology of impurity effects. One of the most important recent developments concerning impurity effects will be discussed in the following section dealing with kinematic theory.

The impurity problem is still far from resolved; for example, little understanding has been gained of the specific mechanisms of interaction between impurities and crystal surfaces or of the marked differences in behavior exhibited by different impurities. However, recent progress has been encouraging and there is reason to expect that important developments will be forthcoming in the near future.

4.2.5. Kinematic Theory of Crystal Growth

One of the problems introduced by the dislocation theory of crystal growth is that of accounting for the velocity and configuration of the steps resulting from the operation of the dislocations. Recently Cabrera and Vermilyea (1958) and Frank (1958B) have developed an important new approach to this problem based on the kinematic wave method of Lighthill and Whitham (1955A, 1955B). By application of this method to a pure system, a "flow concentration" curve is obtained which gives the step flux as a function of step density. For a pure system, the steady state condition corresponds to a uniform step flux, with stability against step

bunching. The effects of transients in the rate of step generation on the step flow and configuration can also be determined.

With impurities present, the flow concentration curve may exhibit inflection points, since the step flux must now depend upon the step density gradient as well as upon step density. This, as well as the time-dependence of impurity adsorption, can lead to the bunching of steps. At present, the analysis has been applied only under conditions where the growth steps are far removed from their sources.

4.2.6. Crystal Dissolution

It is expected that dissolution, as well as growth at singular surfaces, must occur by a dislocation mechanism. However, even at the same driving force, the growth and dissolution velocities are rarely found to coincide. This lack of correspondence is due partly to the different effects of impurities on growth and dissolution, the dissolution rate being much less sensitive to the presence of impurities than the growth process. Frank (1958B) has developed a theory for the dissolution forms of crystals using the kinematic analysis already discussed. The dissolution shape of germanium crystals have been interpreted on the basis of this theory by Ives (1958). However, the subject of dissolution rates and forms resulting from dissolution is one that requires much more extensive and systematic study.

4.3. Growth into Concentrated Fluids

The mechanism of crystal growth (and dissolution) in melts or concentrated solutions is one of the major unresolved problems in the field. As we have seen, the question of the operation of a dislocation mechanism seems to hinge on whether or not the interface is singular. This question has not been clearly resolved on theoretical grounds (Burton, Cabrera and Frank, 1951; Jackson, 1958). However, Cahn (1959) recently has shown that at small undercooling, an energy barrier probably exists, even for a non-singular interface where the crystal-liquid boundary is partially diffuse. If so, growth would occur by the nucleation and growth of steps, but the nucleation process would be much easier than on singular surfaces.

The experimental evidence, based mainly on kinetic and microscopic observations, is still far from conclusive. The evidence seems to indicate that the dislocation mechanism probably does not operate in some systems, e.g., the growth of metal and semiconductor crystals from the melt, but may be operative in others, e.g., the

crystallization of salol, glycerin or PbI_2 from the melt. Hillig (1958) has shown, using Cahn's theory, that some of the kinetic data on nonmetallic materials can be interpreted satisfactorily on the basis that a dislocation mechanism is operating at a partially diffuse (the transition occurring within 2-3 atom layers) crystal-liquid interface.

4.4. Boundary Migration in Crystalline Solids

We summarize here what we believe to be the present state of knowledge concerning the migration of crystal boundaries in crystalline solids. The state of perfection of the crystals formed thereby is treated later. Since the perfection of a crystal will depend, in part, upon the perfection of the grain or nucleus from which it grows, we shall consider first the origin of nuclei for both primary and secondary recrystallization processes.

4.4.1. Nuclei for Recrystallization, Primary and Secondary

Recrystallization nuclei usually stem from one of the following sources:

- (a) Preferred grains
- (b) Inclusions
- (c) Strained domains (i.e., regions within which the dislocation density is much higher than the average).

However, the general or relative importance of each of these sources and the mechanisms whereby they operate are still somewhat obscure.

The conditions under which inclusions or internal preferred grains may function as nuclei have been defined (Turnbull, 1957). Virtually no orientation preferences should exist among nuclei formed from inclusions. Although it is often observed (May, Walter and Dunn, 1959) that primary recrystallized grains do originate at inclusions, the precise mechanism by which nucleation occurs has not been established. That internal grains sometimes serve as recrystallization nuclei is evidenced by the phenomenon of strain-induced boundary migration (Burke and Turnbull, 1952R). Mechanisms whereby recrystallization nuclei might develop from strained domains have been suggested by Cahn (1950) and others. Recently, Hirsch and his associates have observed the development of strained domains into recrystallization nuclei following certain recovery processes (see also Bollmann, 1959).

It is still not established to what extent primary recrystallization nuclei exhibit orientation preferences. However, there is now good

evidence that nuclei for secondary recrystallization quite frequently are grains having particular orientation relations with respect to their surroundings. This orientation preference usually seems to be associated with a lower free energy (these might be, for example, grains which formed earliest in primary recrystallization), or with a lower external surface tension (Assmus, Detert and Ibe, 1957; Walter and Dunn, 1959). Recently, convincing evidence has been obtained by Dunn and Walter (1959) that grains in the outer surface which have the lowest surface tension often become secondary recrystallization nuclei. Furthermore, it appears (Dunn and Walter, 1959) that the relative surface tensions and hence the preferred orientations sometimes can be modified by changing the composition of the surrounding medium.

4.4.2. Crystal Boundary Migration

(a) *Theory.* Dislocation boundaries between slightly mis-oriented crystals migrate upon the application of a suitable stress. This phenomenon (Washburn and Parker, 1952) seems to be fairly well understood (Amelinckx and Dekeyser, 1959R). We shall now consider the migration of boundaries between highly mis-oriented crystals motivated by a reduction in surface or strain energy as the driving force.

The velocity, u , with which such boundaries migrate can be described simply by the equation (Mott, 1948; Turnbull, 1951)

$$u = (D/a) [1 - \exp (\Delta G/kT)], \quad (4.1)$$

where D is a constant having the dimensions of a diffusion coefficient, a is a distance of the order of the lattice spacing, and ΔG is the free energy of activation per atom accompanying the boundary migration. It is assumed that all boundary sites are equivalent.

Much of the early data on boundary migration rates was described by abnormally high apparent activation energies and correspondingly high apparent entropies of activation. Mott believed this implied that boundary migration occurs by a cooperative mechanism, somewhat like local melting, involving many atoms. However, it has been shown (Turnbull, 1951) that impurity effects can lead to abnormally large apparent values of the activation parameters. The results of experiments in which impurity effects have been reasonably well controlled indicate that D in Eq. (4.1) can be identified with the grain boundary diffusion coefficient (Turnbull, 1951) or perhaps even with the liquid state diffusion coefficient (Bolling and Winegard, 1958).

It is expected that D , and therefore u , will depend on the grain boundary structure, which is determined by the relative orientations of the boundary and of the crystals on either side. At the present time, however, a lively controversy exists regarding the exact dependence of u on boundary structure. Liebmann et al. (1956) found that the grain boundary migration rates in aluminum were strongly dependent on misorientation while Graham and Cahn (1956) did not observe such an effect. In view of the recent work of Aust and Rutter (1959) and of Green et al. (unpublished), it appears that the dependence of u on misorientation is not likely to be clarified until the influence of impurities on grain boundary phenomena is better understood.

(b) *Impurity Effects.* It is well established that the migration of crystal boundaries may be greatly retarded or altogether suppressed by a dispersion of impurity particles or by certain soluble impurities. The effects of impurity particle dispersions on crystal boundary migration rates are relatively well understood (Beck, Holzworth, and Sperry, 1949; Zener, 1948; Smith, 1952; Burke and Turnbull, 1952R) on the basis of simple surface tension considerations, but the mechanism of retardation by soluble impurities is still obscure.

Some of the most striking effects produced by soluble impurities have been reported recently by Aust and Rutter (1960). They found that as little as 10^{-6} atom fraction of certain solutes leads to a decrease of a factor of 10 in the velocity of boundary migration in lead, the boundary in this case growing at the expense of a crystal with a pronounced lineage structure. Unfortunately, quantitative data of this type is scarce. From the existing information it appears that the retardation effect increases with decreasing solubility of the impurity.

The usual explanation for this "solute drag" effect (Decker and Harker, 1951; Fullman, 1952; Lücke and Detert, 1957) is that the solute element is positively adsorbed at the grain boundary and that the boundary motion is then limited by the velocity, u_1 , of the solute atmosphere, this presumably being smaller than the velocity of the clean boundary. Decker and Harker, in fact, assumed that u_1 is proportional to the volume diffusion coefficient of the impurity. Using these concepts, Lücke and Detert developed a more detailed theory for the solute drag effect. The formal part of their theory should be approximately correct for the limiting case where the velocity of the solute atmosphere actually is governed by volume diffusion. However, the data indicate that the

solute atmosphere, if it exists, often exhibits a mobility much greater than the *bulk* or *volume* mobility of the impurity.

An additional mechanism of solute retardation is suggested by the recent results of Ainslie et al. (1960) who found that solute atoms diffusing in from grain boundaries can, by the operation of the Kirkendall effect, contribute to the formation of dense dislocation networks on either side of the boundary. Such networks, which are partly stabilized by adsorption of solute atoms, can pin a grain boundary by virtue of the negative gradient in dislocation density on either side of the boundary.

We now consider again the relation between boundary structure and boundary velocity. Aust and Rutter (1960) found little dependence of u on misorientation in ultra high-purity lead. However, the addition of minute quantities of tin led to a strong misorientation dependence; moreover, the migration rates of most boundaries were retarded orders of magnitude more than others separating crystals with certain special misorientations (for example, crystals separated by the Kronberg-Wilson 22° or 38° rotations about a $\langle 111 \rangle$ axis). Green et al. (unpublished) found that misorientation effects in aluminum are prominent only when impurities such as iron or silicon are precipitated from solution. This observation may reconcile the apparently conflicting results reported by Liebmann et al. and by Graham and Cahn. Nevertheless, it is clear that the effects of impurities on grain boundary migration must be carefully described and sorted out before we can arrive at a satisfactory description of the effect of grain boundary structure per se on boundary migration rates.

4.5. Crystal Morphology

Some of the most important morphological problems currently occupying attention are the mechanisms of formation of dendrites, spherulites and whiskers. We will now discuss each of these in turn.

4.5.1. Dendrites

In general, dendrites (tree-like crystals) develop and grow under conditions wherein the driving free energy of transformation is relatively large and the growth is governed by diffusion, e.g., of impurities or of the heat of transformation, away from the crystal surface. It has long been recognized that the dendritic form is one which, because of the "point effect" of diffusion, permits diffusion fluxes much larger than would be possible for polyhedral or spherulitic forms. Thus, it might be expected that

a plane crystal surface, advancing into an undercooled melt or supersaturated solution, would become unstable and would tend to develop protrusions which would quickly grow into a stem as a result of the high diffusion flux. However, the exact manner in which this shape instability originates and develops is presently unresolved. The recent work of Horn (1958; see also Frank, 1958A), indicates that the instability of shape in a diffusion field may be much less severe than previously supposed.

Another important feature of dendrites is that their trunks and branches nearly always are closely parallel to certain simple or low index crystallographic axes. It is understandable that certain axes might be preferred (Chalmers, 1958R; Frank 1958C), but criteria for predicting which axes will be observed under various conditions are still lacking.

Dendrite formation might be a promising area in which to apply the kinematic theory described in section 4.2.5.

4.5.2. Spherulites

One very common morphology is the spherulitic (or its two-dimensional analogue, the cylindritic), in which many crystallites are oriented in a more or less spherical body in such a way that a characteristic crystallographic plane or direction is everywhere approximately tangential to the sphere. It appears that under most conditions polymers crystallize in the spherulitic form. However, this morphology apparently is not limited to complex molecular systems. Thus, spherulitic graphite can form in iron-carbon alloys, and even metal spherulites have been observed (Andrade, 1935). The spherulitic morphology is apparently favored by relatively high driving free energies and kinetic resistances to growth. Basically, however, the mechanisms of spherulite formation are still not understood.

4.5.3. Whiskers

Whiskers are small filamentary single crystals, usually 10^{-6} to 10^{-3} cm. in diameter and up to 1 cm. in length. They have excited great interest because their behavior in some ways approximates that expected of perfect crystals. Some whiskers form by being spontaneously extruded from a surface by what appears to be a recrystallization process (Ellis, Gibbons, and Treuting, 1958) motivated by a reduction in strain energy. Plausible dislocation mechanisms for the extrusion process have been put forth, but these have not yet been substantiated.

Whiskers also form under certain conditions in a supersaturated

medium, most generally a fluid. The supersaturation can be produced as a result of a chemical reaction (as in the formation of whiskers by halide reduction) as well as by the usual physical procedures. In the simplest case, whiskers form in a supersaturated vapor. Here it has been suggested (Sears, 1953) that growth occurs by the following sequence: (a) adsorption of atoms on perfect sidewalls of whiskers; (b) surface diffusion to the whisker tip; and (c) incorporation at a step in the tip associated with a single axial screw dislocation. Gomer (1953) has verified steps (a) and (b) of this mechanism, but the importance of the role played by the dislocation has not been established. Also, no clear explanation of the factors which control the thickness of the whiskers has yet been advanced.

The growth of whiskers in liquids or by the chemical reduction process appears to be more complicated, and there is evidence that impurities sometimes play a decisive role. Price, Vermilyea and Webb (1958) showed that the electrolytic growth of metal whiskers can be interpreted satisfactorily if it is supposed that the sidewall growth is inhibited by adsorbed impurities, and that the rate of tip growth is inherently too great to be interfered with by any time-dependent impurity adsorption. In the chemical reduction process, whisker growth may sometimes be governed by the chemical reaction rate. Plausible mechanisms have been proposed for whisker growth in these circumstances, but none can be considered well established.

5. CONTROL OF PURITY AND STRUCTURE

The key role of impurities in governing the electrical behavior of semiconductor crystals is now well known. Recent research on whiskers and on the growth of large dislocation-free crystals has generated much interest in the control of imperfections in crystals. It is not surprising, therefore, that a very large fraction of the present activity in the crystal growth field is directed toward the control of purity and perfection of crystals. The recent development of high pressure-high temperature techniques has opened up exciting possibilities for the synthesis of new structures, and we may confidently expect a large expansion of research in this area. Essential to research in all these areas are certain basic procedures in crystal growing, which we will now consider.

5.1. Crystal Growing Techniques

It is to be hoped that ultimately, the development of new crystal growing techniques might be based solely upon a fundamental

understanding of the growth processes. In this discussion, we will attempt to highlight those areas in which this ideal has already been closely approached and to point out other promising avenues to the development of crystal growing processes.

5.1.1. General Experimental Problems

The experimental problems can be broadly classified into two main groups: temperature control and control of the chemical environment. For the most part, the experimental problems become critical at elevated temperatures ($> 1500^{\circ}\text{C}$) and pressures.

(a) *Temperature sources and temperature control.* Several reviews¹ dealing with high temperature technique have been published (Campbell, 1956R; Wolfe, 1941R, 1955R).

The ideal temperature source for crystal growing purposes should have the following capabilities or characteristics:

- (1) temperatures in excess of $4,000^{\circ}\text{C}$.
- (2) focusing of the energy in any portion of the system or apparatus without heating undesired regions.
- (3) operation in any chemical environment.
- (4) operation independent of the physical properties of the crystal or its nutrient phases.
- (5) ease of control.

Temperature sources that have been considered for crystal growing facilities are:

- (1) chemical flames.
- (2) electron and ion bombardment.
- (3) "electronic torches" and plasmas.
- (4) radio frequency induction heating.
- (5) radiant sources: radiation-resistance furnaces, solar-type furnaces.

Chemical flames do not meet requirement (3) above. They can be used for direct heating only when the chemical environment of the flame is compatible with the crystal growing system, such as in the growth of corundum crystals by the flame fusion technique. This limitation can be overcome if the flame is separated from the growth system by an impervious refractory shield (indirect heating). Electron and ion bombardment are similarly limited since vacua or reduced pressures are required. Electronic torches (which rely on microwave excitation of diatomic molecules) and plasmas must also be operated in controlled atmospheres.

¹ See also Proceeding of the Symposium on High Temperature (1956; 1958).

Induction heating satisfies all the requirements except (4). It has been widely employed for heating electrical conductors, but cannot generally be used for insulators. At higher frequencies, materials with a wider range of conductivities can be heated inductively, and oxides having specific resistivities of the order of several tenths ohm-cm have been heated at 100 mcps. Experimentation with even higher frequencies could further facilitate the use of this technique.

Radiation-resistance furnaces fail to meet requirements (1), (2) and (3). Metallic resistance heating elements, such as tungsten, may be operated at temperatures near their melting points (3400°C for W) in neutral atmospheres. Unfortunately, no metallic resistor is currently available which gives satisfactory operation in strongly oxidizing atmospheres above 1,500°C, although zirconia- and thoria-base ceramic resistors have been used in oxidizing atmospheres above this temperature. Nevertheless, resistance furnaces for use in oxidizing atmospheres at temperatures above 2,000°C have not been adequately developed.

Focused radiation sources such as solar furnaces satisfy all requirements, except perhaps (5). The principal limitation of solar furnaces are their size and hence their cost, and the lack of reliability of the solar radiation itself. Electric arcs have been substituted for the sun, but they are difficult to control over any lengthy time interval. If a satisfactory primary energy source could be developed, the focused radiation furnace would approach the ideal energy source for crystal growing.

Temperature control requires a sensing device to measure temperature and a feedback mechanism to adjust the energy input. The present limitations of such systems are associated mainly with the temperature sensing element. Here again, sensing devices have been examined thoroughly at temperatures below 1,500°C. The most commonly used devices in this range are thermocouples or resistance thermometers. Thermocouples of specially selected materials can be used at temperatures above 1,500°C, and couples using tungsten, molybdenum and other refractory metals have been developed for operation up to about 2,500°C. There is, however, relatively little experience with these devices, and, at present, each couple requires separate calibration.

Systems utilizing radiation photocells have also been devised for temperature control. Since the emissivity of a solid substance is a sensitive function of the chemical and physical state of the surface, the use of radiation pyrometry must be carefully investigated for each material.

(b) *Control of the chemical environment.* The chemical environment during crystal growing consists of the container or crucible which may be used to hold the growing crystal and the nutrient phases, or both, together with any atmosphere which may be present. The choice of containers is limited by the chemical stability of the container material in contact with the rest of the crystal growing system and by the mechanical properties of the container. As before, the problems in selecting a suitable container material become more critical at higher temperatures. The most commonly used materials are the refractory metals, refractory oxides and graphite.

Among the oxides, fused silica is available in the highest state of purity, is relatively easily fabricated, and is transparent. It is also resistant to thermal shock but cannot be used much above $1,300^{\circ}\text{C}.$, where it devitrifies rapidly and loses most of its mechanical stability. Alumina can be used up to $1,650^{\circ}\text{C}.$ but it is generally of lower purity than silica and is not only more difficult to fabricate but also more susceptible to thermal shock. Stabilized zirconia can be used at temperatures near $2,300^{\circ}\text{C}.$, but fabrication difficulties and thermal shock resistance present problems, as with alumina. Other relatively inert high-temperature materials such as thoria, magnesia, beryllia, boron nitride, and molybdenum disilicide are available, but the ceramic technology of these materials is not yet sufficiently advanced to facilitate widespread application. The development of better high-temperature materials of construction is a problem of special interest and importance to the crystal grower.

Many of the problems of chemical environment can be circumvented by the use of crucible-free techniques which rely on the ability to contain a liquid phase, either by virtue of its own surface tension or with the aid of supporting electromagnetic fields. Exact calculations of the conditions required for the mechanical stability of levitated liquid bodies have been performed for only a few rather idealized geometries. Further theoretical studies in this area might well lead to the discovery of more desirable geometrical configurations. Since this technique, in principle, embodies the ideal conditions for achieving separation of the crystal growing system, both from container materials and from the energy source, it would appear that investigation of the physical principles upon which these levitation techniques are based deserves careful and thorough attention.

5.1.2. Crystal Growing Methods—Requirements and Limitations

General reviews of crystal growing techniques have been contributed by Buckley (1951R), Holden (1950R), Lawson and Nielsen (1958R), Tanenbaum (1959R), Pfann (1957R), and others.² A knowledge of many of the chemical and physical properties of a material is desired for crystal growing purposes. This is because of the extreme structure sensitivity of the growth process, which demands exquisite control of the physical and chemical environment at the growing crystal interface. An attempt will now be made to summarize the various kinds of information that are most critically needed for each particular growth technique or system.

Certain basic information concerning the properties of the crystal is required, this being common to all systems. For example, the density of the crystal and its thermal expansion characteristics are important quantities, since it must be known whether a spontaneously produced nucleus will float or sink to the bottom of the nutrient phase, and whether the grown crystal will be deformed as a result of the stresses set up by differential contraction as the crystal cools to room temperature while surrounded by a crucible. Knowledge of the electrical conductivity of the solid is frequently important in deciding whether electrical heating methods such as induction heating can be used to advantage, and knowledge of the optical properties, such as the emissivity, is required in order to employ radiation techniques for measuring temperature. Knowledge of the thermal conductivity of the crystal and its anisotropy is essential in choosing preferred growth directions and controlling thermal gradients within the growing crystal. In some materials (e.g., bismuth) it is difficult to grow crystals in orientations along which the thermal conductivity is a minimum. In growing alkali halide crystals from the melt, the maximum size is determined by the rate at which the heat of crystallization can be dissipated.

(a) *Growth by liquid-solid transformation.*

(1) *Pure or slightly impure single-component systems.* These systems are often preferred because of their chemical simplicity. If the system truly involves only a single component, mass transport is not required for crystal growth. This permits growth to proceed at rates which are often orders of magnitude faster than the growth rates obtainable in systems which do require mass transport. Moreover, purity control does not pose a serious problem, the shape of the specimen can often be readily controlled, and

² See, for example, Discussions of the Faraday Society on "Crystal Growth" (1949).

self-seeding may generally be accomplished with relative ease. On the other hand, this technique requires the highest temperatures for crystal growth and thus places greater demands upon energy sources and container materials. Furthermore, a single-component system cannot always be employed. In many materials, for example, phase transitions occur below the melting point which effectively destroy a crystal grown at high temperatures.

To grow a crystal from its own pure melt, it is important to know something about the thermodynamic properties of the system as well. These properties include: the melting point; the latent heat of solidification; the vapor pressure at the melting point; the temperatures at which allotropic transformations occur, if any; the free energies of formation of any products which may be formed as a result of possible side reactions with container materials, atmospheres, etc. If the system contains a trace impurity, then the distribution coefficient of this impurity at the melting point should also be determined.

In addition to equilibrium thermodynamic information, kinetic data are also desired. Specifically, it is always helpful if data can be obtained on the thermal conductivity of both the solid and liquid phases, on impurity diffusion coefficients in the solid and liquid phases, on the kinetics of impurity transfer across the solid-liquid interface, on the growth rate anisotropies, and on the kinetics of possible side reactions. Only in rare instances, however, can kinetic data of these types be found in the existing literature.

(2) *Growth of crystals from multicomponent systems.* The growth of a crystal from a solution must take place at a temperature lower than the melting point of the crystal. Lower temperatures are easier to control and maintain, and consequently the container problem becomes less severe. Furthermore, crystals may be grown which have troublesome solid state transformations, or which decompose or melt incongruently. Gray tin and yttrium iron garnet are examples of materials which can be obtained in bulk single crystal form only by growth from solution.

Two major disadvantages of the solution method are that a seed crystal is usually required to avoid excessive nucleation and that the maximum growth rate is much lower than that obtainable in one-component systems. Mass transport of the solute to the growing crystal interface is required and the kinetics of the addition and growth of new layers on the seed crystal are often considerably slower from solution. Finally, with the added pres-

ence of a solvent, the problem of purity becomes increasingly difficult. Gross inclusions of solvent may also occur.

To grow a crystal from solution, certain thermodynamic quantities, in addition to those mentioned above, are of great aid, namely, the solubility of the crystal in the solvent and its temperature coefficient, the vapor pressure of the solvent, and the thermodynamic properties of the solution phase. This last factor is of special importance. If we had sufficient grasp of solution theory to predict solubilities successfully, the empirical techniques which must now be used to find adequate solvents could be eliminated.

Frequently, kinetic information of various kinds may prove to be of even greater value than thermodynamic data. The most useful kinetic information is that dealing with the kinetics of solvation and other possible side reactions involving solute and solvent, the kinetics of solvent evaporation, the viscosity of the solution, the diffusion coefficient of the solute, and growth rate anisotropies and the influence of impurities thereon. Problems raised by impurities are of special importance, since the key to many of the mysteries of crystal growth probably lies in the adsorption of impurities on the surfaces of the growing crystal. Studies of adsorption from liquid phases on inorganic solids would contribute greatly to an understanding of growth processes.

(b) *Growth by vapor-solid transformation.* Some materials sublime to the vapor without melting. If no adequate solvent is available, crystals of such materials may sometimes be grown by a vapor-solid transformation. Growth rates are usually intermediate between those achieved in one-component systems and in growth from solution. The problems which exist are similar to those encountered in growing crystals from solution. Difficulties arise with regard to the control of growth rate anisotropies and the suppression of spurious nucleation and, moreover, long times are required to grow large crystals.

In addition to the usual thermodynamic data, knowledge of the vapor pressure and its temperature coefficient is required. Information concerning the chemical reactivity of the vapors with any carrier gases that may be used, or with the containers in which crystal growth is proceeding, must also be available. Kinetic data, such as the rates of molecular and thermal transport in the vapor phase are required in order to predict the maximum growth rate. Whenever the molecular species in the vapor differs from that in the solid (as with silicon carbide), a surface barrier may exist to vaporization, and the kinetics of the vaporization process

itself may be strongly influenced. The chemical rearrangement required at the surface may thus determine the rate at which a crystal can be grown.

(c) *Crystal growth above atmospheric pressure.* Pressure is as significant a variable as temperature in crystal growing processes, and important structure syntheses (e.g., diamond and quartz) can be achieved successfully by the application of high pressures. The biggest obstacle to high pressure crystal growth is the difficulty of finding corrosion resistant materials which can withstand high stresses at high temperatures. The development of new materials of construction which will permit the operating temperature range to be raised by a factor of two and the pressure range by an order of magnitude would completely revolutionize the field. Improved design and increased availability of high pressure vessels would also stimulate more research in this field. Until it becomes easier for the average institution to obtain, from commercial sources, high pressure vessels having a wider range of usefulness than presently available, experimental studies of crystal growth at high pressures will be somewhat restricted.

The same kinds of thermodynamic data are needed as are required for crystal production at atmospheric pressure, except that these data must now be extended to include pressure as one of the thermodynamic variables. Until recently, the high pressure field has been virtually neglected and thermodynamic data are scarce, but in order to take advantage of high pressure techniques, a fairly detailed knowledge of the equations of state of both solid and liquid phases is essential.

Existing information concerning kinetic processes at elevated pressures is even more fragmentary. The pressure coefficients of viscosity, diffusion, and thermal conductivity are seldom available and it appears that it will be necessary to develop new experimental techniques to obtain such data. Techniques which are familiar at ordinary pressures such as optical, electrical, and thermal conductivity measurements, electropotential determinations, etc., must all be adapted for high pressure use (Jamieson and Lawson, 1959R). Even methods for the accurate measurement of extreme pressures are currently unsatisfactory and require further development; see Part XII, section 3.

Another subject which deserves explicit mention is the phenomenon of the metastability of thermodynamically unstable phases. It is known that in many systems metastable phases may crystallize out of a liquid or vapor phase in preference to the thermodynamically stable phase (e.g., the polymorphism of SiC).

It is also commonly observed that some metastable phases may exist for indefinitely long periods because the rate of transformation to a more stable phase is extremely slow (e.g., diamond). Both of these phenomena may frequently be traced to the nature and kinetics of the nucleation process. This specific problem has received very little attention but is of considerable practical importance to the crystal grower.

(d) *Growth of crystals by chemical reaction.* Crystals are sometimes produced by carefully controlling the rate at which a chemical or electrolytic reaction occurs, the product of the reaction being the material from which the crystal is grown. This method can sometimes be used to grow crystals when no adequate solvent can be found.

The practical application of this method requires simultaneous control of the following steps:

1. the diffusion of reactants to the reaction site
2. the rate of reaction
3. the addition of the desired reaction product to the growing crystal
4. the diffusion of any undesired reaction products away from the reaction zone, and
5. the dissipation of both the heat of reaction and the heat of crystallization.

Conditions under which the rate-controlling steps are compatible with the growth of a single crystal must be determined. This is perhaps the most complex of all the crystal growing processes discussed here. In addition to the kinds of information required for other processes, knowledge of the chemical equilibrium and the rate constants of both the desired reaction and any possible side reactions is also required, together with an understanding of the heterogeneous nature of these reactions.

5.2. The Control and Determination of Impurity Concentrations in Solids

General reviews on this subject have been published by Kröger and Vink (1956R) and by Hannay (1959R). The need to control the concentrations of impurities in solids is becoming increasingly important. Control implies not only purification but also the ability to introduce specific impurities into the solid phase in definite, known concentrations (Thurmond, 1959R). In many cases, as with donors and acceptors in semiconductors, these concentrations must be made to vary from one region to the other within a single crystal in a definite, prescribed manner.

There are many aspects to the problem of impurity control and at least five can be identified as current problem areas. These are: 1) techniques for impurity determination; 2) phase equilibria; 3) the thermodynamic properties of dilute solid solutions; 4) the kinetics of transport of components within the liquid, gas and solid phases and between these phases; 5) the interactions between impurities and other lattice defects.

5.2.1. Impurity Determination

A principal difficulty in controlling impurities and studying their effects is the task of analyzing for the impurity. Some property of the solid phase can usually be found which is sensitive to the impurity and can thus be used to determine the concentration of the impurity, but only after calibration by chemical analysis. Frequently, it is the identity of the impurity itself which must be determined and chemical analysis again is usually the only method available. As a result, the development of more universal and rapid methods of microanalysis will continue to be of special interest to solid state scientists.³

The following analytical techniques appear to be particularly useful and their continued development is therefore a matter of considerable urgency or importance: Activation analysis, when applicable, is a powerful technique although it is tedious and requires highly specialized equipment. Mass spectrometric techniques for the analysis of solids show great promise but further work is needed in order to develop and perfect the volatilizing procedures. Emission spectroscopy has been extended into the microrange for certain impurities. Electrochemical methods such as coulometry and polarography are useful for certain analyses, and solution absorption spectroscopy, which makes use of various complexing agents has proven to be a valuable microanalytic technique.

The determination of the lighter elements, such as boron, nitrogen and phosphorus, continues to be difficult. The selective analysis of micro quantities of the rare earth elements also constitutes a difficult problem, although ion exchange techniques show significant promise.

³ A more extended discussion of the capabilities and limitations of existing techniques for the analysis of trace impurities, and techniques which hold promise for future development, is presented in Part XII; see also Inghram (1959R).

5.2.2. The Equilibrium Diagram

A knowledge of the phase diagram is of primary importance whenever the transport of components between phases is to be studied or exploited. A detailed phase diagram for a multi-component system may be quite complex and difficult to determine. The starting point, however, is the binary phase diagram.

There are relatively few binary phase diagrams which are completely known. Hansen (1958R) in his compilation of binary phase diagrams gives at least some information on a total of 1380 out of the 3180 binary combinations possible with the 65 elements he has considered. Of these 1380 systems, partial or complete diagrams are given for 750 systems, but of this group there are probably less than a dozen for which the solid solubility of each component in the other is reasonably well known over a significantly wide temperature range. In many systems, upper solubility limits have been estimated, but, while giving a certain amount of helpful information, these estimates are hardly sufficient in situations where careful control must be exercised.

There is a dearth of information on slightly soluble impurities. Oxygen, nitrogen and carbon are of particular interest because they are often a major source of contamination. Experimental data on the solubility of other slightly soluble impurities are also highly desirable because of the fundamental thermodynamic information which can be inferred from such measurements.

Of particular interest in impurity control is the distribution coefficient of an impurity near the melting point of the host solid (Pfann, 1957R; Thurmond, 1959R). The distribution coefficient, k , is defined as the ratio of the atom fraction of the impurity in the solid phase to its atom fraction in the liquid phase, i.e., $k = x_s/x_L$. This ratio gives important information relative to the ease with which the impurity concentration can be reduced by fractional crystallization or zone refining. It also specifies the concentration of impurity that must be added during growth from the melt in order to obtain a desired concentration or impurity level in the grown crystal. The distribution coefficient can be evaluated from the liquidus and solidus curves in the region of the melting point of the host solid. As stated above, this information is seldom available, especially for slightly soluble impurities. The measurement of the distribution coefficient in dilute solutions near the melting point frequently provides a great deal of information about the complete solid solubility curve.

5.2.3. Thermodynamic Properties

One goal of those interested in impurity control in solids is the prediction of solid solubilities or distribution coefficients. This goal could be achieved if it were possible to calculate, over a wide range of concentrations, the thermodynamic functions (i.e., the partial molal free energies, heats and entropies) of the impurity in the liquid and solid phases. For slightly soluble impurities (small k), a calculation of the thermodynamic properties of dilute solid solutions would probably suffice. The study of dilute solutions presents a reasonable challenge to the theoretician because either the concentration dependence of the interactions can be neglected or, if not, then this dependence can be accounted for with relative ease. The study of solid solutions offers certain advantages over liquid solutions because of our better general understanding of the structure of the solid state. Thus, studies of slightly soluble impurities in solids could provide a major contribution to the theory of solutions.

The estimation of the thermodynamic properties of liquid solutions at high temperatures is also of interest. Most of these estimates will undoubtedly continue to be derived from experimental correlations but it is hoped that theory will offer some guidance. The interest here arises principally because liquid phase interactions also have an influence on distribution coefficients. Furthermore, the solubility of the major component is related to the thermodynamic properties of the liquid phase. The ability to estimate these thermodynamic properties becomes an important problem in the interpretation of small distribution coefficients.

There is evidence, from the behavior of impurities in germanium and in silicon, that the distribution coefficients for a series of impurities at temperatures near the melting point can be correlated with the heat of vaporization of the pure solute element. Correlations of this type permit empirical interpolations and extrapolations which enable estimates to be made for impurities having unknown distribution coefficients in germanium and silicon. Studies of the fundamental significance of such empirical correlations may permit similar predictions to be made in other systems. At present, few measurements of k are available for materials other than germanium and silicon. Measurements of the distribution coefficients for a series of impurities in other host solids would be welcomed, particularly for systems of solid state interest such as the alkali halides, transition metal oxides, etc. Data of this type are essential for the growth of crystals with specified properties.

5.2.4. Kinetics

The diffusion rates and mechanisms of diffusion of impurities in solids are also of vital concern to those interested in impurity control. In addition to affecting the distribution of impurities that are incorporated into a crystal during growth from the liquid phase, solid state diffusion may also be employed to introduce impurities into grown crystals. Interest also attaches to the diffusion of impurities in the phases from which the crystals are grown. While thermodynamic considerations determine the equilibrium distribution coefficients of impurities, kinetic processes determine the effective distribution coefficients that pertain during the growth of a real crystal. The diffusion of rejected impurities away from the interface ($k < 1$), or the diffusion of impurity into the depleted region at the interface ($k > 1$), are important in determining the effective distribution coefficient.

The kinetics of the process responsible for the transport of impurity across a phase boundary are also of fundamental interest in determining the effective distribution coefficient. Practically nothing is known about such interface reactions and any studies which will provide information relative to the rate constants and the mechanisms for these processes will be pertinent.

Finally, convective mixing and stirring in the liquid phase are kinetic processes which are not sufficiently well understood at present to permit satisfactory control of the composition of the interface layer, and thus to control the incorporation of impurities into crystals grown from liquid phases.

5.2.5. Interactions of Impurities with Other Lattice Defects

Lattice defects such as vacancies, interstitials, electrons, holes, or other chemical impurities may significantly alter the optical, electrical, magnetic and thermodynamic properties of a solid solution. The interaction between an impurity atom and a vacancy, an electron or hole, or another impurity, either substitutional or interstitial, may be largely coulombic or covalent, or it may best be described in terms of the interactions between stress fields. The distribution of impurity atoms in a solid solution, and hence the properties of the crystal, will depend on the concentrations of various lattice defects which are present, on the nature and magnitude of the interactions between impurity atoms and other defects, and on the thermodynamic equilibria and kinetics governing these interactions (Kröger and Vink, 1956R).

When the interaction is primarily coulombic, its magnitude may be calculated if an effective charge and a dielectric constant

can be assigned. For short-range coulombic interactions, the designation of the dielectric constant offers particular difficulties. Covalent interactions are poorly understood. Estimates of the magnitude of these interactions, based upon the interactions which occur in elemental or compound semiconductors, may possibly be helpful in this connection. Estimates have been made of the lattice relaxation in the vicinity of a simple vacancy and of stress fields around impurities but the choice of an appropriate radius for the impurity, and of the elastic constants, frequently poses difficulties, particularly if the impurity is in interstitial solution.

At least at elevated temperatures, the concentrations of various impurity-defect complexes are frequently governed by the requirement that homogeneous chemical equilibria be established between the various defect species present in solid solution. Typical examples are the association equilibria involving electrons and holes with donors or acceptors in silicon and germanium (Fuller, 1959R) and vacancy-divalent impurity interactions in the alkali halides (Watkins, 1959). The equilibrium constants for defect reactions of this type, and their temperature coefficients, are usually not available or are known only approximately.

5.3. Perfection Control

The mechanical, physical, and chemical properties of crystalline materials are influenced to varying degrees by crystalline imperfections (Shockley et al., 1952R; Fisher et al., 1956R). Generally, crystalline imperfections can be divided into three groups: gross imperfections, line defects, and point defects. This discussion will concentrate primarily on line defects, that is, dislocations.

Although gross imperfections can normally be eliminated by growing a single crystal using standard techniques, nearly all single crystals contain dislocations in varying concentrations up to roughly 10^{12} per sq. cm. Unfortunately, many experiments are still being conducted on materials of non-reproducible and, in fact, unknown dislocation structure, this despite the known influence of dislocations on some properties or a total lack of information concerning the structure-sensitivity of other properties. The degree of perfection currently obtainable varies considerably, depending upon the material. At one extreme, silicon, germanium, silver chloride, and probably ice have been obtained in the form of bulk single crystals that are completely dislocation free, while soft metals are not yet available with dislocation densities much below about 10^6 cm⁻².

There are two facets to this general problem. One is the avail-

ability of experimental techniques and the necessary background of knowledge concerning the principles involved in the *control* of crystalline perfection. Perhaps more crucial, however, is the present inadequacy of experimental methods for the *measurement* of crystalline perfection. Therefore, it becomes important to consider experimental methods for the detection and determination of defects in crystals (Amelinckx, 1959R).

Because of current interest in their structure, their unusual mechanical properties, and their growth mechanisms, micro-crystals and whiskers will be discussed in some detail. It has been established that some of these crystals, particularly whiskers, often display the remarkable strength expected from dislocation-free material, but there is substantial evidence that whiskers do, in fact, frequently contain some dislocations.

5.3.1. Determination of Dislocation Structure

The availability of appropriate methods for the determination of dislocation structure constitutes one of the major problems associated with the control of crystal perfection. Such techniques should be nondestructive, applicable to crystals containing a wide range of imperfections, and capable of revealing the finest details of the dislocation structure in the entire volume of relatively large crystals without the necessity of cutting the crystal. Unfortunately, none of the methods currently available satisfies all these criteria.

Gross estimates of the perfection of single crystals can be obtained in a variety of ways. The most common method relies on observation of the angular distribution of diffracted X-rays, as in the Laue, Berg-Barrett, Guinier-Tennevin and Schultz techniques, all of which depend upon misorientations of lattice planes within the crystal to reveal the presence of inhomogeneous distributions of dislocations in clusters or sub-boundaries (Barrett, 1952R; Guinier and Tennevin, 1949). Another simple method, which is applicable to crystals which exhibit prominent cleavage, is the observation of cleavage surfaces. If cleavage is properly carried out, the degree of smoothness of the cleavage surface may be related directly to the crystalline perfection of the material. This, of course, is only a qualitative method. In certain cases, an estimate of perfection can be obtained by measurement of various bulk properties, such as the electrical resistivity, thermoelectric power, and, for semiconducting materials, the carrier lifetimes. However, these properties are also highly sensitive to the presence of point defects and impurities, and consequently other informa-

tion about the chemical and structural perfection of the crystal is necessary for a proper analysis.

More detailed methods which identify individual dislocation sites are also available, but these can generally be applied only to crystals of relatively high perfection. These techniques include: (a) etch pit methods (Lovell, Vogel and Wernick, 1959); (b) decoration with precipitates (Barber, Harvey and Mitchell, 1957; Amelinckx, 1957, 1959R); (c) utilization of X-ray diffraction intensity contrast and extinction (Lang, 1958, 1959; Newkirk, 1958; Webb, 1958); (d) electron transmission contrast (Hirsch, Horne and Whelan, 1956; Hirsch, 1959R; Whelan, 1959); (e) anomalous X-ray transmission (Hunter, 1958); and (f) field ionization microscopy (Müller, 1958). We will now consider these methods individually; see also Part XII, section 2.

(a) *Etch pit techniques.* Techniques of this kind have been developed for many materials. Any etch pit technique for the identification of dislocations must be proven or established using the particular material in question, since etch pits are often developed at sites other than the intersections of dislocations with surfaces. The method, of course, reveals only the dislocation structure at the surface. It is nondestructive, except that the surface is somewhat altered by the etching procedure. In principle, the dislocation structure in the interior of a large crystal can be determined by etch pit techniques simply by sectioning and carrying out etch pit studies on a collection of sectioned surfaces. The principal difficulty is that the sectioning must be accomplished without introducing additional dislocations into the crystal. It appears that the development of etch pits is controlled by the selective action of impurities which inhibit crystal dissolution elsewhere, but less so near dislocations. However, the phenomenon is still not well understood. A better understanding might make it possible to eliminate much of the trial-and-error effort currently needed in order to develop suitable dislocation etchants, and to apply this simple and highly convenient, though incomplete, technique to virtually any material of interest.

(b) *Decoration techniques.* These methods are destructive since they consist of a heat-treating procedure which usually results in an irreversible precipitation or segregation of a solute along dislocation lines. In transparent materials, the decoration can be observed directly using an optical microscope (Amelinckx, 1957, 1959R). For silicon, which is optically opaque, infrared illumination has been used to reveal decorated dislocations (Dash, 1957). Decoration techniques have been applied effectively to the alkali

halides, the silver halides, silicon, germanium, and sapphire, and, with the aid of etch pit techniques, to copper and iron-silicon alloys.

(c) *X-ray diffraction intensity contrast and extinction techniques.* This method is based on the fact that relatively imperfect regions near dislocation cores diffract X-rays more strongly than the relatively perfect regions remote from dislocations. It has been applied to the determination of the dislocation structure of silicon, germanium, lithium fluoride, aluminum, sodium chloride, and zinc (Lang, 1959). The technique suffers from the limitation that only a thin layer of the order of magnitude of the mean penetration depth of the available characteristic X-rays can be examined. Unfortunately, the development of this technique is probably being retarded by a lack of understanding of the basic theory of extinction contrast in the diffraction of X-rays from crystals containing dislocations.

(d) *Electron transmission contrast technique.* Contrast in the transmission and absorption of electron beams has been used most effectively to study the dislocation structure of thin films (Hirsch, 1959R). This technique is discussed in more detail in section 7.2.

(e) *Anomalous X-ray transmission technique.* This new technique is based on the phenomenon of anomalous transmission of X-rays through nearly perfect crystals at the Bragg angle. In principle, it should be possible to map the dislocation structure of relatively perfect crystals by exploitation of this phenomenon. The technique already has been used to obtain a general estimate of crystal perfection (Webb, 1959).

(f) *Field ionization microscopy technique.* This is the only presently known method for studying the details of the atomic configurations in the neighborhood of dislocations. By applying to a crystal surface an electrostatic field of the order of 10^8 volts/cm., it is possible to ionize an ambient gas in such a way that the pattern of the ion beam produced on a phosphorescent screen corresponds to the atomic configuration of the edges of lattice planes which comprise the surface. The resolution is of the order of 2\AA and a magnification on the order of one million-fold can be obtained. The specimens are required to be in the form of electrically conductive needles having a tip radius of approximately 1000\AA ; the surface area available for study in a single specimen is therefore only on the order of 10^{-10} cm^2 . This area is a factor of 10^4 less than the area per dislocation in a normally imperfect metallic specimen, hence special techniques are required to provide appropriate samples containing dislocations; see also Part IX.

The size of specimens is sufficiently small that the observed behavior of dislocations may not correspond precisely to the behavior of dislocations in bulk crystals. No extraneous thermal treatments are required to clean the specimen surface, thus making it possible to study the effect of thermal treatment unambiguously. Edge dislocations, stacking faults, slip bands, and possibly screw dislocations have been observed. Although the detailed atomic configuration around dislocations can be studied using this technique, the interpretation is subject to two limitations: 1) only the intersection of dislocations with surfaces is observed, and 2) while under observation, the surface is always subjected to stresses of 10^{10} to 10^{12} dynes/cm², these being near the theoretical strength of most materials. Nevertheless, this technique may properly be considered the highest resolution technique currently available for solid-state studies of atomic configurations.

5.3.2. Control of Dislocation Structure

(a) *Bulk Crystals.* In general, it is possible to estimate, on the basis of experience, the degree of crystalline perfection that would be obtained in a particular material using the more or less conventional techniques employed in the current art of crystal growth. Usually, the harder the material the lower the attainable dislocation density. The ranges of dislocation densities currently attainable for different classes of materials are tabulated below:

Class of Materials	Dislocation Density (cm ⁻²)
Hard insulators	0 to 10^5
Hard semiconductors	0 to 10^5
Semimetals	10^5 to 10^6
Hard metals	10^5 to 10^6
Ductile metals	10^6 to 10^8

Perhaps the most notable achievement has been the production of completely dislocation-free crystals of silicon (Dash, 1958) and germanium as large as 2 cm. in diameter and 30 cm. long.

It appears that the attainable perfection depends more upon the type of material involved than on the particular growth method employed. On the other hand, recently discovered recipes for the production of dislocation-free crystals indicate that dramatic improvements in the standard techniques of crystal growth are certainly possible.

(b) *Microcrystals and Whiskers.* Comprehensive reviews on this subject are given by Brenner and by Nabarro and Jackson in GPC (1958R). The most naive considerations suggest that

microcrystals might be nearly dislocation free if they exhibit the normal dislocation densities characteristic of bulk crystals. Thus, if bulk metal crystals normally contain a dislocation density of the order of 10^6 per sq. cm., then there should be a substantial probability that a crystal on the order of 10^{-2} cm. in maximum dimension would contain no dislocations or perhaps only one. Indeed, this idea seemed to be confirmed with the discovery that some whiskers display almost the full theoretical strength expected for completely dislocation-free materials. However, further experimental work has indicated that whiskers frequently do contain dislocations.

Whiskers are often grown under conditions which strongly suggest that dislocation mechanisms are operative in their growth. Therefore, it would be expected that at least one screw dislocation would necessarily be present in each of these crystals. This is, however, not inconsistent with the high observed strength, since a single axial screw dislocation would have little influence on the strength properties. X-ray diffraction studies have revealed the presence of one or more dislocations with a screw component along the axis of whiskers grown from many materials. In most metal whiskers, on the other hand, no axial screw dislocations have been found (GPC, 1958R), contrary to expectation.

Recently, experiments on whiskers and platelets of sodium chloride (Amelinckx, 1958; Webb, unpublished) have indicated that these crystals may contain several dislocations parallel with their growth axis without detriment to their strength, whereas other whiskers, containing only a few dislocations, are weak, apparently because the particular configuration of dislocations present provides dislocation mills that are active at low stresses. Small areas of surface damage have also been found to be particularly deleterious to strength.

Observed dislocation densities in sodium chloride whiskers are actually greater than those found in bulk crystals. Although sodium chloride is the only material for which the dislocation structure of whisker-like microcrystals has been determined in complete detail, and correlated with observations of strength, there is mounting evidence for the belief that many microcrystals actually contain normal or greater than normal dislocation densities, and that the observed mechanical properties of these crystals depend not upon this gross dislocation density but, rather, on the particular configuration of dislocations that exists. The dislocation configurations that are present in most metal whiskers have not yet been determined unequivocally.

Other microcrystals besides whiskers may sometimes be expected to contain only a few dislocations. For example, cadmium iodide crystals grown from solution have been observed to grow by the Frank mechanism in a way which indicates that they probably contain only a few dislocations (Newkirk, 1955). Some platelets of zinc grown from the vapor phase have exceptionally high strength and show no evidence for dislocations when examined using techniques available at present; other platelets clearly show grown-in edge dislocations and have very low strength (Webb, unpublished).

Our understanding of the dislocation structures of microcrystals is complicated somewhat by certain other considerations. Thus, if it requires a higher stress to move than to generate dislocations in the material, then the strength of the crystal should be unaffected by the presence or absence of dislocations. Indeed, it has been demonstrated that in certain materials, such as silicon, the introduction of dislocations has no effect on the room-temperature strength (Pearson, Read and Feldman, 1957). Size effects also become a problem in considering the mechanical properties of extremely small crystals. The critical Griffith crack size is $\sim 1 \mu$ and the critical dislocation loop size is approximately 500 Å for stresses of the order of the theoretical stress in most materials. As the specimen dimensions are reduced to this size range, drastic changes in behavior are to be expected and such changes have, in fact, been observed. For example, in metal films roughly 500 Å thick it has been shown that although dislocations are present, they are often immobile, and as a result failure occurs without appreciable plastic deformation (Menter and Pashley, 1959R).

(c) *Special Case: Dislocation-Free Crystals.* The unlimited number of possible dislocation structures which could exist in imperfect crystals suggest that the only configuration which would be completely reproducible would be the dislocation-free structure. Fortunately, it is possible to obtain such structures. Dislocation-free silicon and germanium have been grown from the melt using the Czochralski technique (Dash, 1959). Silver chloride single crystal platelets, several centimeters square and a fraction of a millimeter thick, have been grown essentially free of dislocations by recrystallization (Mitchell, 1958). Hence, at least two techniques are *potentially* available for the production of dislocation-free crystals of other materials, but the general extent to which these techniques may be applicable has not yet been established.

(d) *Mechanisms for Dislocation Generation During Growth.* The primary mechanisms responsible for the generation of dis-

locations during crystal growth are still not well established. There are only a few recognized possibilities and it is unlikely that any additional, fundamentally different mechanisms will appear. Dislocations may originate in the seed or parent phase and be propagated by the growth of the crystal. They may originate at an internal or surface defect, or at composition inhomogeneities involving a lattice parameter change, any of which may provide a source for easy slip or a source of stress concentrations. Dislocations may also be formed by the occlusion of growth fronts with imperfect lattice matching. During the cooling of the crystal, they may be generated by the collapse of vacancy platelets or clusters formed by the condensation of individual vacancies.

Subsequent to the operation of any of these mechanisms, substantial dislocation multiplication may occur under stresses which result from differential thermal contraction during growth or cooling, or from accidental loading produced by improper handling procedures. If a crystal is grown dislocation-free and if dislocation sources are not introduced subsequently by vacancy condensation, etc., then thermal stresses are probably not too important, although sufficiently high local stresses can, of course, nucleate dislocations even in a perfect crystal.

The operation of a dislocation mechanism during the growth of a crystal should inevitably result in a dislocation structure. Thus, the fact that silicon and germanium crystals have been grown from the melt without dislocations indicates that, at least for these materials, a dislocation mechanism is not essential to crystallization from the melt. Although it has not been established unequivocally, it seems probable that dislocations are normally essential for the growth of large crystals from solution. If so, we would not expect to be able to grow dislocation-free crystals of most materials from solution.

The growth of dislocation-free crystals of silver chloride by a recrystallization technique suggests that this method may perhaps be generally adaptable to the production of dislocation-free crystals of other materials. In the simplest model, a growing grain might be expected to drag along its own dislocations which terminate in the boundary, while destroying the dislocations in the adjacent grain that is being consumed. If these were the only processes which occurred, virtually perfect crystals could always be grown, in the manner Mitchell used for silver chloride, by arranging for the grain boundary to fan out or increase its area as it moves. This process would be limited only by generation of dislocations by the moving boundary or by the passage of disloca-

tions through the boundary. Unfortunately, it appears from the work of Dunn (1959) on silicon-iron and of Berghezan (unpublished) that the moving boundary does sometimes act as a source of dislocations. Nevertheless, there is a strong possibility that the dislocations might have been generated by the passage of the boundary around impurity particles, such as slag or oxide inclusions. If so, perfect crystals might be obtained by careful prior purification of the material.

The behavior and generation of dislocations during the growth process are governed by certain characteristic properties of dislocations in crystals. There are, for example, topological restrictions on dislocation geometry and these in turn suggest methods for controlling the generation of dislocations. Moreover, the variation of the energy of the dislocations with orientation provides the key to the development of techniques by means of which the propagation of dislocations during growth from an imperfect seed crystal can be prevented. Favored crystallographic directions of propagation are observed in silicon and germanium, and the growth directions make the largest possible angle with the growth direction, thus forcing the dislocations out through lateral surfaces during growth (Dash, 1959).

Although the existence of preferred dislocation directions can be understood, at least partially, on the basis of lattice geometry, the problem is actually one of much greater scope and should preferably be approached from the viewpoint of crystal chemistry as well. In the past, the energetics of dislocations have usually been ignored possibly because dislocations are non-equilibrium defects. It is felt, however, that a true understanding of the pronounced variations in dislocation behavior which are observed among different crystalline materials will probably not be achieved until relationships are established between the energetics and mobility of individual dislocations and the nature of the binding forces in crystals. Another aspect of this problem which has resisted attempts at understanding is the stabilization of dislocation structures by the formation of networks and by solute segregation. These problems recur in many parts of solid state physics but are particularly relevant to crystal growth and to the mechanical behavior of crystalline solids.

5.4. Synthesis of Substances into Different Crystal Structures

An important use to which crystal chemical knowledge can be put is the prediction of structures. Thus, using crystal chemical principles it is possible to deduce that some metallic oxides of

composition MO_2 will have the fluorite structure and others the rutile structure, depending on the radius of the metallic ion. Accordingly, crystal chemistry serves as a guide for the synthesis of new substances with desired structures. Quite often it is the only guide, since pertinent thermodynamic data are rarely available.

The extent to which this guidance is helpful depends on the nature of the substances under consideration and the state of development of their crystal chemistry. A high degree of confidence can be justified for ionic substances, where the principles are well-established and rather satisfactorily obeyed. Two recent examples of considerable complexity may be cited: (1) the ferri-magnetic rare-earth garnets, illustrating both the role of ionic size in determining which rare-earth ions may be incorporated in the structure (Geller, 1960) and the rule that a pair of ions (Si^{4+} and Ca^{2+}) may be replaced by a different pair (Fe^{3+} and Mn^{3+}) with the same total charge, and (2) synthetic micas (Wones and Eugster, unpublished) in which large concentrations of iron are substituted for aluminum and silicon.

Structures can also be predicted quite successfully in crystals with covalent binding, as attested by the large number of zinc blende structures which have recently been synthesized. On the other hand, the prediction of metal structures, where the crystal chemistry is less certain, is in general not very reliable, although it may be quite good in special circumstances. When dealing with two elements which differ considerably in size, and when geometrical factors are dominant, the reliability of prediction of an alloy structure may be reasonably high. The large number of Laves phases and the extensive list of compounds isomorphous with NaZn_{13} bear out this point.

It should be emphasized that very often the synthesis of new materials is undertaken because of a strong technological motivation. Thus, the need for special magnetic materials has provided the major impetus for the development of the rare-earth garnets. These materials could have been predicted a long time ago and might have been synthesized much earlier had sufficient motivation existed. The same is true concerning the synthesis of new zinc blende-type structures, which has been stimulated by the need for better semiconducting materials.

Methods and techniques of synthesis are mainly outside the scope of crystal chemistry, but high pressure techniques, intelligently used in conjunction with crystal chemical considerations, may be singled out as an especially attractive means of obtaining

new materials. The general criterion is that if crystal chemical principles indicate the possible existence of a structure of higher density than the structure which is stable at atmospheric pressure, then high pressure would favor its formation. Substances with structures of lower coordination than indicated by the radius ratio (or with a radius ratio slightly less than the critical value for phase transition) might bear examination in this respect. Thus, the possibility of synthesizing CaO and other alkaline earth oxides in the CsCl-type structure might be considered. Another possibility is that the layered CdCl_2 - or CdI_2 -type structures may transform at high pressures to the denser rutile structure. These are merely simple illustrations of the numerous possible high-pressure syntheses that can be suggested by the application of crystal chemical principles.

Another way in which it is possible to obtain substances in a metastable structure is by a nucleation and growth process whereby nucleation of the more stable phase is avoided. This is readily accomplished using systems in which crystal growth is sluggish since, under these conditions, the relatively high supersaturations necessary for the growth of metastable phases can be maintained for long periods. Recently, it was discovered that metastable phases can sometimes be produced in the form of droplets under conditions where the crystal growth rates are actually very high (Cech, 1956). This phenomenon may be traced to the effective isolation of nucleating impurity particles which is achieved by dispersing the specimen into droplets. The potential value of this droplet technique for structure synthesis has by no means been fully exploited.

Still more recently, an ingenious procedure has been developed by which small amounts of liquid alloys can be quenched rapidly enough to forestall the normal nucleation and growth processes (Duwez, Willens, and Klement, 1960). This technique has already led to the synthesis of new metastable phases in a number of alloy systems and promises to yield important information on the stability of alloys.

Metastable structures (e.g., supersaturated solid solutions or metastable crystal structures) may sometimes be produced by electrodeposition processes, which can be carried out at such low temperatures that the rate of formation of the more stable structure is negligible.

6. CRYSTALLIZATION OF POLYMERS

Fundamental solid state studies on polymers have in the past

been hindered seriously by the complex mode of polymer crystallization and the unavailability of polymer single crystals. If the complex phenomenology of polymer crystallization could be unraveled and controlled, we are confident that the basic solid state science of crystalline polymers would be a rich area for exploration. Many of the topics touched on here are treated in more detail by Flory (1953R), Bunn (1953R), Mandelkern (1956R, 1958R), Keller (1958R), and Stein (1958).

6.1. Introduction

Despite the fact that, in the molten or amorphous state, polymer molecules possess a large amount of configurational freedom, chains which are sufficiently regular in structure can undergo partial three-dimensional ordering under appropriate thermodynamic conditions. This process is termed crystallization, and although equilibrium considerations lead to the conclusion that very high degrees of crystallinity should be attained in homopolymers, this condition is not usually fulfilled because of kinetic factors involved in the transformation. Thus, a polycrystalline system evolves, the degree of crystallinity and the resulting morphology determining to a large extent the physical and mechanical properties of the material.

Though equilibrium conditions are difficult to attain in a crystalline polymer, it has been shown that the transformation between the crystalline and liquid state corresponds to a first-order phase transition and that the appropriate laws governing phase equilibria are applicable. This conclusion has an important bearing on the analysis of the crystallization mechanism and morphology, and places the problem within the appropriate scientific framework. In considering the crystallization properties of macromolecules, the various types of chain molecules available and their modes of crystallization must be clearly distinguished. For example, crystallinity can be developed by various techniques so that a preferential crystallite orientation exists or, alternatively, the crystallites can be arranged in a random manner. The crystallization process can be expected to be highly sensitive to the chain structure, and particularly to the presence of structural irregularities incorporated within the chain, the chain length, and the distribution of these quantities. Before conclusions of general validity can be drawn from experiment, these structural factors must be carefully assessed.

6.2. Structure and Morphology

A vast quantity of experimental evidence has been accumulated which indicates that organized structures exist in polymers at all levels of morphology accessible to study. Unit cell structures, as determined by wide-angle X-ray analysis, indicate that bond angles, bond distances, and other elements of structure are closely similar to the corresponding quantities deduced for low molecular weight materials. It is not uncommon to find more than one chain passing through the unit cell and various types of chain configurations, ranging from planar zig-zag to helical arrangements, are observed in crystals of various polymers. It can be concluded that the repeating chain unit plays a role analogous to that of molecules in monomeric organic crystals.

At higher levels of morphology, polymer molecules assert themselves in somewhat unique fashion and, despite intensive study, no unified, generally accepted picture of the morphology has as yet emerged. In axially-oriented polymers, several orders of discrete, low-angle X-ray diffraction maxima, directed along the meridian, are observed, indicating a well-developed periodicity along the fiber axis. The spacings, which range in magnitude from 50 Å to 500 Å, can be interpreted as resulting from the distribution of crystalline and amorphous material along the fiber axis.

Low-angle diffraction maxima are also observed in non-oriented but highly crystalline polymers. In such materials, the spacings are not preferentially directed, but depend markedly on the crystallization temperature and are also of the order of several hundred Å. The temperature dependence of the spacings suggests that the crystallization process may be controlled by a nucleation mechanism. However, no quantitative explanation of the low-angle diffraction maxima has yet been developed. At a higher level of morphology, studies of the angular distribution of light scattered by thin films of crystalline polymers indicates an organized structure of dimensions comparable with the wavelength of light. It has been concluded that the scattering is due to structures which involve crystallites.

When viewed under a light microscope, a still higher level organization of crystalline material is observed. This manifests itself in the form of spherulitic structures which, though not unique to polymers, represent a common if not a universal mode of homopolymer crystallization. Various types of spherulitic structures have been observed, depending on the polymer type and crystallization conditions. Despite the intensive study that spherulite formation has received, certain basic problems are still un-

answered. Although the spherulitic structure is frequently observed, it is as yet not conclusively established that it does, in fact, constitute a universal mode of crystallization from the melt. A great deal of descriptive material exists concerning the nature of spherulites, but the reason why they form remains obscure, nor has the molecular mechanism involved been established.

When polymer molecules are crystallized from dilute solutions, electron microscope observations indicate that plate-like crystals are obtained which consist of thin lamella about 100Å thick. Thickening occurs by a screw dislocation mechanism, and electron diffraction studies have clearly identified the platelets as single crystals. These recent experimental observations indicate that single crystals of polymers can actually be synthesized, although the maximum size currently obtainable is such that the crystals are visible only under the electron microscope. No doubt further progress in the techniques of growing large single crystals can be achieved; toward this end, detailed studies of the crystallization mechanism would be most helpful.

In studying the properties of the single crystals, the significant observation has been made that the chain axis of the molecules is oriented normal to the flat faces of the platelets. Low-angle X-ray diffraction data also indicate a well-defined periodicity, the spacings varying from about 100 to 200Å depending on the crystallization conditions. Since the extended length of a typical molecule is of the order of thousands of angstroms, the conclusion has been reached that a periodic and systematic folding of the molecule occurs during crystallization. This accounts rather simply for the orientation effects that are observed as well as for the height of the steps which are manifest on the faces of the platelets. However, a satisfactory explanation, based on fundamental considerations, has not yet been offered for why chain folding should be favored either kinetically or thermodynamically. Other possible chain arrangements in the crystal consistent with the morphological observations should perhaps be considered. A fundamental problem is therefore raised regarding the structure of the nucleus for crystallization from dilute solution as well as from the melt. Recently, electron micrographic evidence has been presented which suggests that chain folding is the basic mode of crystallization in bulk polymers.

6.3. Mechanism of Liquid-to-Crystal Transformation

In an effort to achieve a better understanding of the crystallization mechanism, a great deal of attention has been focused on

answered. Although the spherulitic structure is frequently observed, it is as yet not conclusively established that it does, in fact, constitute a universal mode of crystallization from the melt. A great deal of descriptive material exists concerning the nature of spherulites, but the reason why they form remains obscure, nor has the molecular mechanism involved been established.

When polymer molecules are crystallized from dilute solutions, electron microscope observations indicate that plate-like crystals are obtained which consist of thin lamella about 100Å thick. Thickening occurs by a screw dislocation mechanism, and electron diffraction studies have clearly identified the platelets as single crystals. These recent experimental observations indicate that single crystals of polymers can actually be synthesized, although the maximum size currently obtainable is such that the crystals are visible only under the electron microscope. No doubt further progress in the techniques of growing large single crystals can be achieved; toward this end, detailed studies of the crystallization mechanism would be most helpful.

In studying the properties of the single crystals, the significant observation has been made that the chain axis of the molecules is oriented normal to the flat faces of the platelets. Low-angle X-ray diffraction data also indicate a well-defined periodicity, the spacings varying from about 100 to 200Å depending on the crystallization conditions. Since the extended length of a typical molecule is of the order of thousands of angstroms, the conclusion has been reached that a periodic and systematic folding of the molecule occurs during crystallization. This accounts rather simply for the orientation effects that are observed as well as for the height of the steps which are manifest on the faces of the platelets. However, a satisfactory explanation, based on fundamental considerations, has not yet been offered for why chain folding should be favored either kinetically or thermodynamically. Other possible chain arrangements in the crystal consistent with the morphological observations should perhaps be considered. A fundamental problem is therefore raised regarding the structure of the nucleus for crystallization from dilute solution as well as from the melt. Recently, electron micrographic evidence has been presented which suggests that chain folding is the basic mode of crystallization in bulk polymers.

6.3. Mechanism of Liquid-to-Crystal Transformation

In an effort to achieve a better understanding of the crystallization mechanism, a great deal of attention has been focused on

kinetic studies of the transformation of bulk homopolymers. Analysis of the temperature coefficient of the rate of isothermal development of crystallinity from the melt, as well as the rate of spherulite formation at low values of undercooling, indicates that a primary nucleation mechanism is operative, since the observed temperature coefficients are in accord with nucleation theory as appropriately modified for polymer systems. Whether primary nucleation occurs homogeneously or whether it is always heterogeneous has not been clearly resolved and experimental evidence for both points of view is available. To settle this question, experiments should be performed similar to those carried out in metal systems, wherein the crystallizing substance is isolated in the form of small droplets; in addition, a more detailed thermodynamic analysis of the solid-liquid interface should be made. As indicated previously, another fundamental problem concerns the shape of and disposition of the chain elements within the nucleus.

Analysis of the growth rate is hampered by the inherent difficulty of observing and measuring the size of the fundamental crystallizing entity. Existing information on the growth process has been deduced, therefore, primarily from studies of the spherulitic growth rate, but it should be kept in mind that crystallite growth and spherulitic growth are not necessarily identical processes. It is found that the radius of the spherulites increases linearly with time, and that the temperature dependence of the growth rate is characteristic of a nucleation-controlled reaction. Two types of nucleation processes have been postulated. One is the classical two-dimensional nucleation event and the other a three-dimensional nucleation process initiated or stimulated by the primary nuclei already formed. The latter process would be unique to polymer systems. Although, in principle, these two nucleation processes should be distinguishable on the basis of the functional dependence of the growth rate on undercooling, the diffuseness of melting does not allow for the measurement of the degree of undercooling with sufficient accuracy.

Kinetic studies have also been made on the rates of polymer crystallization from very dilute solutions, i.e., in the composition range where single crystal platelets are formed. Surprisingly, the isotherm shape, temperature coefficient and crystallization rate constants are remarkably similar to those observed for bulk homopolymers, indicating that the same general mechanisms are operative.

In summary, kinetic studies have led to certain general conclusions in regard to the crystallization mechanism. It seems clear

that nucleation and growth processes are involved and strong evidence exists that the growth is also nucleation-controlled. With the establishment of these general principles, much remains to be done in order to elucidate, in a more quantitative manner, the details of the nucleation and growth processes, their inter-relations and their time dependencies. It should be emphasized that in its present form, the theory of the crystallization kinetics of polymeric systems is still very rudimentary, and care should be taken that the basic assumptions which have been made are recognized when interpreting experimental results.

Although the crystallization mechanism and the morphology of flexible chain molecules have received the most consideration to date, another class of macromolecules—the fibrous proteins—occur naturally in the crystalline state with highly developed axial orientations. The individual polymer molecules in these substances are very asymmetric, and the problem is to explain the mechanism by which the individual molecules crystallize so as to develop the organized fibrous structure which is observed. Theoretical and experimental work on this problem is still in its infancy and though no general overall pattern has as yet emerged, the principles of crystallization by nucleation and growth would still be expected to be applicable.

7. FORMATION AND STRUCTURE OF THIN FILMS

The properties exhibited by crystalline bodies, especially the structure-sensitive properties, are often unusual whenever the physical size of the body becomes very small in at least one dimension. We have already discussed the unique behavior of droplets and whiskers which are small in three and two dimensions, respectively. Now we take up the subject of the formation and structure of crystal bodies which are small in just one dimension, that is, thin films (Neugebauer, Newkirk and Vermilyea, 1959R). It is believed that the unique properties of thin films, although partly due to the shape anisotropy per se, largely have their origin in special microstructural features which arise during the formation and growth of the films. In the discussion which follows, therefore, considerable emphasis is placed on the microstructural features of thin films.

7.1. Production of Reproducible Substrates for Film Growth

The difficulties in specifying and controlling the structure and chemical composition of solid surfaces have already been emphasized. However, these difficulties must be surmounted in some way

if results that contribute significantly to the understanding of film growth mechanisms are to be obtained.

One important method of investigating surface structure is through the use of refraction effects in reflection diffraction photographs. This technique gives some information on surface smoothness and facet development but not much additional detail. High-resolution replica methods are now capable of revealing surface steps of the order of 10\AA in height (Bradley, 1959) but even greater sensitivity is required in order to reveal every atomic step on the surface. A valuable technique developed by Bassett (1958) which appears to do just this, at least in some situations, is that of decorating the surface steps with small particles of vacuum-evaporated gold. At present, the technique is limited to ionic crystal surfaces and further development or refinement of the technique is needed before it may be extended to other systems.

The ideal surface on which to grow or deposit a thin film would be a clean, molecularly smooth, single crystal surface. Four techniques are commonly used for the preparation of good single crystal surfaces, namely:

1. Direct growth of crystals from the vapor or from solution.
2. Electropolishing
3. Vacuum evaporation
4. Cleavage

Of these, the cleavage method would seem to offer the most promise for obtaining a surface with characteristics approaching the ideal. Ordinarily, however, surface steps are produced during cleavage. These steps result from dislocations having components of their Burgers vector normal to the cleavage surface and (Gilman, 1956; Forty, 1957) from plastic deformation during cleavage, particularly when the rate of propagation of the crack is low. It would undoubtedly be of value to study the cleavage behavior of the best or most nearly perfect crystals now obtainable at high rates of crack propagation. In this way, more control over the surface geometry might be obtained. It appears that the most reliable procedure presently available for eliminating indeterminate surface contamination is cleavage in ultra-high vacuum.

A monolayer of oxygen has been found sufficient to inhibit the epitaxial growth of copper on titanium (Schlier and Farnsworth, 1958). It is important to know, therefore, whether the surface presents, to a deposited film, an array of atoms similar to that observed on a section through the crystal parallel to the surface, or whether the surface consists of an array of completely different atoms, e.g., oxygen. For this and other reasons, further

studies of residual adsorbed gas films using the technique of diffraction of slow electrons would be most helpful.

Single crystal films formed by oriented growth on single crystal surfaces are in themselves good substrates for the subsequent growth of a thin film of a different material. Pashley (1959) has found that the outer surface of a silver film $\sim 2000\text{\AA}$ in thickness, grown epitaxially upon a mica cleavage surface, is not only extremely smooth but also uniform in orientation, with the [111] direction normal to the surface, as determined by reflection electron diffraction. This technique has many attractive features as a method for obtaining reproducible single crystal metal surfaces, and it would be valuable to extend it to other metals, if possible.

Before terminating this brief discussion of substrates, it should be noted that single crystals in the form of thin plates or ribbons have been prepared unattached to substrates, either by vapor decomposition of metal halides on metals or by growth from solution (some alkali halides). Detached crystals of this type present some difficulties in handling, but there may be situations in which they are useful; consequently, it would be of interest to investigate in more detail the growth conditions required for the development of exaggerated platy habits in crystals which normally show an equiaxed habit.

7.2. Characterization of Defects in Thin Films by Electron Microscopy

It has been common practice in the past to attempt to characterize the defect structure of thin films by observations or measurements of various macroscopic properties. Until recently, the nearest approach to a direct method was the technique of transmission electron diffraction. By the use of a microbeam or highly selected-area technique, it is possible from measurements of the number of spots on diffraction rings, and the size, shape and displacement of these spots, to deduce useful information concerning "particle" size, misorientations, sub-boundaries between particles, elastic strains, stacking faults, etc. (Hirsch, Kelley and Menter, 1955). This approach is, however, extremely tedious and the information obtained is not precise. A revolution in technique has occurred in the last few years arising from the application of the high-resolution electron microscope to this problem. There are now three distinct ways of determining defect structures by direct observation:

- (1) The microscope is used near its resolution limit in an effort to resolve lattice periodicities directly and thereby detect the

presence of departures from perfect periodicity which may be associated with imperfections (Menter, 1956, 1958R). At present, this technique appears to be limited to a resolution of about 5Å so that it has been possible to observe dislocations only in molecular crystals with rather large periodicities. From the point of view of general application, particularly in relation to the possibility of conducting experiments on the specimen while it is under observation, e.g., by changing stress, temperature, etc., the technique of transmission electron microscopy is to be preferred over the field ion emission method of imaging atomic arrays (sec. 5.3.1.). In order to extract more information concerning the detailed configuration of atoms around dislocation lines, there is a need for considerable theoretical work, using the dynamical theory of electron diffraction, on the origin of contrast in images of crystal lattices. Periodic images of a type similar to the directly resolved crystal lattice may also be produced by superlattices. New and detailed information relating to periodic antiphase structures has already been obtained, for example, in the copper-gold system by Pashley and co-workers (Glossop and Pashley, 1959; Pashley and Presland, 1959). The relation between antiphase boundaries and dislocations is also amenable to study. There is clearly an immense field to be explored here which was formerly only accessible by the averaging techniques of X-ray diffraction.

(2) The problem of resolving periodicities smaller than those directly accessible with present electron microscopes has to a large extent been overcome by the application of moiré techniques, involving the formation of a periodic pattern by the superposition of two crystals. Bassett, Menter and Pashley (1958) have shown how this technique may be applied in a controlled manner to obtain a pattern whose geometry is directly related to that of the two superposed crystals. Dislocations in either crystal as well as elastic lattice distortions can be observed by means of moiré patterns.

(3) In the third type of investigation, structural defects in the film are made visible as a result of abnormal electron scattering in their neighborhood. This method has been extensively developed by Hirsch and his co-workers (for a recent review, see Hirsch, 1959R) in the last few years and detailed theories have been developed of the contrast effects associated with dislocation lines and stacking faults (Whelan, 1959).

To summarize, it is now possible by a variety of techniques to observe and characterize planar and line defects in thin films by transmission electron microscopy. As yet, no direct evidence has

come forward that individual point defects can be detected, although it is likely that quite small aggregates or clusters of such defects will ultimately become visible. If the defect cluster is sufficiently large, of course, details such as dislocation loops arising from collapsed vacancy discs can be observed (Hirsch and Silcox, 1958).

7.3. Oriented Overgrowth and Growth of Single Crystal Films

The phenomenon of oriented overgrowth (epitaxy) has attracted the attention of crystallographers for a great many years. Oriented deposits may be formed on substrates either by vacuum evaporation, vapor plating, electrodeposition, deposition from solution, or by chemical growth. Epitaxial growth by vacuum evaporation represents one of the best methods for growing thin single crystal films suitable for a variety of fundamental studies. We shall consider now the nucleation and growth of such films (Bassett, Menter and Pashley, 1959R).

Despite all of the work that has been done in this field, particularly by optical microscopy and electron diffraction, it remains true that we still cannot define precisely the conditions which lead to epitaxial growth. Surface cleanliness and smoothness of the substrate are important variables and it is frequently advantageous to work at elevated temperatures, inasmuch as this may improve the surface cleanliness and also increase the surface mobility of the deposit atoms. Similarity in lattice symmetry is desirable, but not essential, and the original assumption of Royer that similarity of bond type was desirable has been shown to be relatively unimportant. The main concern or emphasis of many past investigations has been in establishing possible correlations between the orientations and misfits of the substrate and overgrowth lattices. A recent summary of the situation by Pashley (1956R) shows, however, that misfit alone is not an adequate guide.

In considering nucleation mechanisms, it is important to distinguish, on experimental grounds, between whether growth starts by the spreading of a uniform monolayer of the deposit material over the surface, or by a process involving the formation of discrete "island" nuclei. It has been shown that reflection electron diffraction techniques are sufficiently sensitive to detect monolayer films if the substrate is sufficiently smooth; indeed, a number of cases have been observed where this type of growth probably occurs. It appears, however, that the vast majority of overgrowths are produced by the second process, i.e., by the growth of distinct

nuclei. This suggests that in most cases the equilibrium form of the initial deposit is determined by a balance between the surface energy of the deposit, the interfacial energy between deposit and substrate (this will include a term for the energy of the immobile dislocation net at the interface, in addition to the energy arising from the normal chemical interaction between the substrate and overgrowth) and the internal energy of the deposit, which may also include a component of strain energy arising from the misfit at the interface. It is therefore likely that most experimental situations are more complex than that postulated in the Frank and van der Merwe (1949) model for oriented overgrowth. The problem is in many respects geometrically similar to that encountered in the formation of Widmanstätten precipitates in the interior of crystals. A theoretical treatment of the problem along these lines was given by Turnbull and Vonnegut (1952).

Electron microscope studies of thin foils of aluminum-copper alloys by Nicholson and Nutting (1958) have provided evidence for the existence of the coherency strains between precipitates and matrix in the early stages of precipitation. It should be possible, using thin single crystal films as substrates, to detect elastic strains which may exist between the substrate and the nuclei of an overgrowth formed upon such a substrate. There does not appear to be any technical reason why the whole experiment of depositing an overgrowth by evaporation on to a substrate maintained at an elevated temperature could not be carried out inside the electron microscope, and observed directly in transmission.

Another important factor in the nucleation of overgrowths is the possibility of chemical reaction at the interface, including the possibility of solution of one phase in the other. Glossop and Pashley (1959) have found that in some metal/metal systems alloying can occur in the initial stages of deposition when there is some mutual solubility of the two metals. Alloying is observed in some systems even at room temperature where bulk diffusion rates are small. The process is probably aided or accelerated by the high vacancy concentrations which are likely to be "quenched" into the film during growth. This phenomenon is of fundamental importance both from the theoretical and practical point of view, since, on the one hand, it will undoubtedly affect the interfacial energy between substrate and overgrowth, and on the other, the adhesion of the film will be modified.

Singularities on the substrate surface also may play an important role in the nucleation of films. Optical studies have revealed that macroscopic steps are often preferred sites for nucleation,

and more recently Bassett (1958) has established that monatomic steps may sometimes act as preferential sites for nucleation, e.g., in the deposition of gold on alkali halides. Evidence that homogeneous nucleation may occur at points removed from surface steps was also obtained by Bassett, although the possibility cannot be excluded that this nucleation is also determined by surface singularities such as point defects or adsorbed impurity atoms. The possible role of dislocations, i.e., surface steps associated with dislocations having a Burgers vector component normal to the surface, has been discussed many times. Here again, the application of thin film transmission techniques should be most rewarding. Dislocations in the film are readily characterized so that it should be possible to obtain direct visual evidence for any correspondence between dislocations and nuclei of a second phase. It is particularly important to determine the extent to which the defect structure of a substrate influences the defect structure of a film grown upon it.

7.4. Microstructure of Single Crystal Films

The growth of single crystal films by vacuum evaporation on to single crystal substrates must certainly be regarded as one of the most remarkable developments which have appeared during the last few years. By the careful choice of the growth conditions, it is possible to grow continuous, substantially parallel-sided metal films in single crystal orientation, the areas obtained being limited only by the area of the substrate. Although no measurements of density have been made on films of this type, visual inspection of their microstructure would suggest that, in marked contrast with the behavior of fine-grained films, the density is not substantially different from the bulk density. The films can be made available either attached to a substrate or they can be detached, if the chemistry of the system permits solution of the substrate without attacking the film.

Most work so far has been carried out on gold films in (111) orientation formed on silver/mica substrates (Pashley, 1959). Films with a (100) orientation may be grown either by direct evaporation on to a rocksalt cleavage face or on to a silver/rocksalt substrate. At least with the gold/silver system, it appears that by electropolishing any arbitrary face on a silver single crystal, and using this crystal as a substrate, a gold film may be grown in parallel orientation with the substrate; in other words, a quite general method is available for growing single crystal films of gold in any orientation. Much more work will be needed

in order to develop suitably compatible systems for other materials, and efforts in this direction would be highly desirable.

Detailed study of gold films has shown that they have a high dislocation density of the order 10^{10} to 10^{11} per cm^2 . When the films are thin, the dislocation lines are isolated. They terminate on the surfaces of the film and generally follow the shortest path possible on the slip plane from one surface to the other. As the film thickness increases beyond about $1,000\text{\AA}$, the dominating influence of the surface in keeping the dislocation lines straight tends to diminish, and, at the limit of observation determined by the transparency of the film to electrons (about $2,000\text{\AA}$), a network of dislocation lines seems to be developing. More information is required about the changes in dislocation structure with thickness.

Some films, particularly those grown on rocksalt, contain many twins and stacking faults, the origin of which is not clear. All films appear to contain a fine-scale structure, some of which can be resolved into small dislocation loops less than 100\AA in diameter. At least in part, this structure may be associated with the aggregation of vacancies during film growth (Bassett, Menter, and Pashley, 1959R), but the additional role of impurities, particularly from the gas phase, cannot be ignored. All of these features undoubtedly have their origin in the nucleation and growth process. A thorough investigation of the manner in which the structure develops will be of extremely wide interest, since crystal growth by this technique of building up a thin film, atom-by-atom, is one which is susceptible to direct and detailed observation while it is actually occurring. The results obtained from these studies may therefore be of general value in understanding the processes whereby imperfections are built into crystals grown by other techniques.

7.5. Thin Film Properties

The successful production of thin, single crystal films, together with the development of sensitive, high-resolution techniques for studying their microstructure, have laid the foundations for a comprehensive investigation of the relation between microstructure and properties of thin films. The results obtained from studies of this kind on single crystal films will be of inestimable value in interpreting the properties of films prepared under less ideal conditions. The situation is analogous to that encountered in the study of materials in bulk, in that the properties of polycrystalline aggregates cannot be properly understood without a detailed

knowledge of the behavior of bulk single crystals. We shall indicate briefly here the types of investigation contemplated.

(1) Detailed study is required of the influence of temperature on the defect structure, since this is basic to a number of other properties. We need to know, for example, how effectively the dislocation density can be reduced by annihilation and/or climb out of the film. The behavior of individual point defects and clusters of point defects is also of importance in this connection.

(2) Systematic studies are required in order to establish the effect of the rate of formation of the film and of the residual gas pressure on the defect structure.

(3) Measurements of mechanical strength (Pashley, 1959; Menter and Pashley, 1959R) have already shown that factors other than film thickness, most probably associated with defects in the film, exert a controlling influence. These factors must be understood before the behavior of polycrystalline films can be fully explained.

(4) Observations on internal stresses in thin films and on the relationship between residual stresses and microstructure are needed. These are particularly important since the use of thin films in many applications is limited by the tendency of the films to crack as a result of internal stress. We are considering here stresses which are distinct in origin and superimposed upon those arising from differences in the expansion coefficients of film and substrate. Explanations of their origin have been given which are based upon the behavior of surface forces in polycrystalline aggregates. The elimination of internal boundaries by the use of continuous single crystal films should simplify the study of these residual stresses.

(5) Direct correlations should be sought between resistivity changes and microstructural changes which occur on ageing and annealing.

(6) Attention should be given to the production of good single crystal films of magnetic and dielectric materials and studies should be made of the properties of such films in relation to structure.

(7) The dependence of microstructure on thickness should be studied on detail. In some cases, orientation is said to improve with thickness whereas, in others, the reverse appears to be true. Care is needed in this type of experiment to ensure that the environmental variables do not change with thickness; for example, the radiant heating of a substrate by an unscreened source during a prolonged evaporation is not negligible. Thicknesses out-

side the range of the transmission electron microscope may be handled by using some appropriate thinning procedure (Kelly and Nutting, 1959). A better procedure might be to use the high-voltage electron microscope to avoid the need for thinning. The application of this instrument to structural problems has, however, not been explored very intensely as yet.

(8) Reliable methods should be developed for preparing single crystal alloy films of predetermined composition. The ability to make such films in precise, simple orientations would greatly facilitate the study of alloys, since thin films produced from a bulk specimen of arbitrary orientation would undoubtedly exhibit complex structures that would be most difficult to interpret. Some alloys may also be expected to have microstructures differing from those of the bulk, and these may turn out to have valuable properties.

8. SUMMARY OF CONCLUSIONS AND CURRENT IMPORTANT PROBLEM AREAS

We shall now attempt to summarize our conclusions and our assessments of the present frontiers of knowledge in the field, emphasizing the important problem areas of current interest. This has its dangers and limitations since only the more obvious or well-defined frontiers can be easily recognized. It would not be surprising, in fact, if some of the important future discoveries in the field were predicated on the recognition of problems that are not quite so obvious today.

8.1. Crystal Chemistry

The generalizations of crystal chemistry have been applied quite successfully in explaining and correlating the crystal structures of ionic and covalent substances. Although far from infallible, these more-or-less empirical generalizations serve as the most useful guide now at our disposal for synthesizing ionically and covalently bonded materials into new structures. On the other hand, the application of crystal chemistry to the interpretation of metallic structures has been somewhat less successful. As yet, a corresponding set of crystal chemical generalizations does not even exist for molecular or polymeric crystals.

Important Questions and Problem Areas:

(1) *Justification of crystal chemical generalizations at a more fundamental level.* The present status of this problem is discussed in some detail in Part II of this volume, which deals specifically with the cohesive properties of solids.

(2) *Are there a few simple principles by which packing in molecular crystals is governed and, if so, how can these principles be formulated?*

(3) *What are the packing principles which govern the structures of polymeric crystals?* Here a need may exist for the development of special stereochemical principles.

(4) *Elucidation of the structure of non-crystalline condensed phases, i.e., liquids, liquid crystals and glasses.* In this connection it appears that the study of liquid crystals has been somewhat neglected in the past and that it offers a promising field for future investigation. The present state of knowledge concerning the structure and properties of liquids is reviewed at some length in Part X of this volume.

8.2. Theory of Growth and Morphology

The fundamental problem here is to characterize precisely the nature and structure of crystal interfaces and the mechanisms of various atomic interactions which occur at these interfaces. Apparently, we now have a fairly good understanding of the structure and mechanism of movement of crystal-dilute fluid interfaces. However, our present knowledge concerning other types of interfaces, except those between slightly misoriented crystals of the same composition, is much more primitive. Important advances have recently been made in the phenomenological description of impurity effects in crystal growth, but we still lack a satisfactory understanding of the specific role of impurities in most situations. *Important Questions and Problem Areas:*

(1) *What is the detailed atomic structure and mechanism of movement of the interface between a crystal and its own melt or between a crystal and a concentrated solution of its components?* The solution to this problem probably hinges, in part, on an understanding of the structure of the liquid state.

(2) *What is the structure and mechanism of migration of highly misoriented crystal-crystal boundaries?* The solution to this problem would undoubtedly help to clarify the mechanism of recrystallization and the origin of recrystallization textures.

(3) *What are the specific roles of impurities in crystal growth processes?*

(4) *Problems in morphology, including the origin of spherulites, the mechanism of dendrite formation and the mechanisms of whisker growth.*

8.3. Control of Purity and Structure

8.3.1. Crystal Growing Techniques

Crystal growing procedures have reached a fairly high state of development. Provided the pertinent material constants (kinetic and thermodynamic) are known, it is almost always possible to design a satisfactory crystal growing procedure for a particular substance. The procedure which generally yields the purest and most perfect crystals is growth from the melt, but this method is not universally applicable, often because of the intervention of solid state transformations or undesirable side reactions. Unfortunately, the material constants needed for designing crystal growth procedures are frequently not available. Aside from this, the main problem in the field at present would seem to be the extension of growth techniques to very high temperatures and pressures.

Important Problem Areas:

(1) *Development of techniques for growing crystals at high temperatures, i.e., above 1,500°C, both at atmospheric pressure and at very high pressures.* The main requirements for these developments are:

- a. Improved methods for generating high temperatures and pressures
- b. Better methods for temperature measurements and control
- c. Materials of construction which will withstand extremely high temperatures and pressures

(2) *Determination of basic material constants.* It would be out of the question, of course, to undertake an experimental program designed to provide a complete set of kinetic and thermodynamic constants for every material. However, when single crystals of a particular class of materials are required, an effort should be made to determine the pertinent constants for that class of materials.

8.3.2. The Control and Determination of Impurity Concentrations in Solids

The techniques for controlling impurity concentrations in solids are now quite well developed. Here again satisfactory procedures usually can be worked out if the required material constants are known, but difficulties arise in certain systems because of the inadequacies of present methods of analysis for trace impurities. At a more fundamental level, satisfactory theories or generaliza-

tions concerning the thermodynamic properties of dilute solid solutions are still lacking.

Important Questions and Problem Areas:

- (1) *Determination of basic material constants.*
- (2) *Are there some simple generalizations which describe the thermodynamic behavior of dilute solid solutions?*
- (3) *What are the detailed atomic interactions which occur between impurity atoms and other lattice defects?*

8.3.3. Perfection Control

Within recent years, several excellent methods have been developed for revealing the positions of dislocations within crystals or their points of emergence at the surface. Through the use of these techniques, dislocation theories have been confirmed in broad outline, although many questions of detail remain. Activity stimulated by the success of dislocation theory has led, among other things, to the recognition of the unusual properties of crystal bodies which are physically small in one or more dimensions, and to the synthesis of bulk dislocation-free crystals of some materials. Present indications are that the exceptional properties of whiskers are due to unusual dislocation configuration rather than to the complete absence of dislocations per se. Much of the future progress in this field hinges upon achieving further refinements in the techniques for revealing dislocations and other defects in crystals. These developments, in turn, may depend upon advances in our understanding of the phenomena on which the experimental techniques are based.

Important Questions and Problem Areas:

- (1) *Methods for characterizing dislocation structures.* Despite recent progress in this area, much remains to be done before we are able to determine the dislocation structure of any given substance in any arbitrary physical form. Two of the most promising methods, namely, X-ray diffraction contrast and etch pit techniques, might be greatly advanced by the development of more complete or detailed theories of the phenomena on which they are based.
- (2) *What is the origin of dislocations during the growth and thermal treatment of crystals?* We know relatively little about how dislocations may be generated by a moving boundary; for example, whether or not impurities are essential to the process is far from clear at the present time. Dislocations often form subsequent to the growth process, and although plausible mechanisms

have been suggested by which this can occur, the details of these mechanisms are still quite obscure.

(3) *What is the reason for the stability of certain dislocation configurations in crystals?* After a large number of dislocations have been introduced into various crystals it is usually impossible to reduce the dislocation density below a certain level, e.g., 10^5 to 10^6 cm⁻², by any practicable annealing treatment. Why this is so, and whether or not impurities play a decisive role are questions which remain unanswered at present.

(4) *How do the properties of dislocations vary with the structure and type of binding of the crystal (i.e., what is the "crystal chemistry" of dislocations)?*

(5) *Development of the art of growing perfect crystals.*

(6) *What is the relation between the dislocation structure and the properties of whiskers?*

8.3.4. Synthesis of Substances into Different Crystal Structures

It appears that crystal chemical principles offer perhaps the best guide for exploring the possibilities of new crystal structure syntheses. New structures may be synthesized in a variety of ways, most commonly by chemical reaction, by the application of high pressures, or by controlled nucleation and growth. The general applicability of the high pressure method has been greatly extended by recent developments in techniques for achieving and maintaining very high pressures at high temperatures.

Important Question and Problem Area:

(1) *What further possibilities for new structure synthesis are indicated by crystal chemistry?* With the new techniques of synthesis now available, a more systematic look at this problem might be worthwhile.

8.4. Crystallization of Polymers

The crystallization of polymers, like the crystallization of other materials, is a first-order phase transition which occurs by nucleation and growth. However, because of the complexity of the molecular structure, the morphology of crystalline polymers is generally quite complex, and the crystallization process usually does not go to completion. Furthermore, the growth of polymer crystals seems to be governed by a secondary nucleation process which occurs periodically. As yet, procedures have not been developed for growing large single crystals of polymers. However, the success recently obtained in growing small isolated polymer crystals constitutes an extremely important and noteworthy achievement.

Important Questions and Problem Areas:

(1) *What is the mechanism of nucleation?* It is still not established to what extent nucleation in bulk polymer systems is homogeneous, and a satisfactory theory for the crystal-melt interface does not exist.

(2) *What is the mechanism of growth?* Although it seems clearly established that growth is controlled by a secondary nucleation process, particularly at small undercoolings, the specific mechanism is still in doubt. In addition, the role played by transport processes at large undercoolings and the relation of the growth kinetics to glass formation is in need of more quantitative study.

(3) *Growth and structure of single crystals.* In order to facilitate the development of single crystal growing techniques, it is important that the nature of the chain disposition within the crystals and the crystallization mechanism from dilute solutions be firmly established. It is also necessary to define more completely the conditions under which chain folding will or will not occur during crystallization.

(4) *What is the origin of the spherulitic structure?* Basically, it would be desirable to establish the mechanism of formation of spherulites in polymers, and whether any relation exists between the mechanism operative in polymers and that governing the formation of spherulites in nonpolymeric systems. Furthermore, the arrangement of chain units within polymeric spherulites is still unresolved. More rapid progress toward the solution of these problems might conceivably be made if morphological and kinetic studies on spherulites were carried out simultaneously.

(5) *What effect does the molecular weight and distribution of weights have on the crystallization kinetics?* It is important that cognizance be given, in future studies, to the possible effect of molecular weight and distribution, and of the need for careful specification of the crystallization conditions, particularly the undercooling at which transformation occurs. This will, in turn, require rather careful thermodynamic determination of the melting temperature, the importance of which has not been fully realized in morphological studies. The influence of non-crystallizing chain irregularities on the nucleation and growth processes should also be carefully assessed.

8.5. Formation and Structure of Thin Films

Techniques for investigating the structure of thin films have been developing rapidly, and, as a consequence, we now have con-

siderable information about defect configurations in thin films. Moreover, methods have been developed for growing thin films in the form of single crystals, but as yet these techniques have been adapted only to a comparatively small number of substances. Thin films, like other bodies which are small in one or more dimensions, often exhibit unique physical properties or behavior. It appears that this uniqueness, although partly due to shape anisotropy per se, has its origin largely in special microstructural features which arise during the formation of the films.

Important Questions and Problem Areas:

- (1) *What is the mechanism of nucleation in thin films?* Fundamental information is lacking not only about nucleation under ideal conditions, but also about the role of surface defects and adsorbed impurities in both the nucleation and growth processes.
- (2) *What is the mechanism of formation of defect structures in thin films and how are these structures modified as the film thickens?*
- (3) *How are the properties of thin films related to their defect structure?*
- (4) *How can the techniques for growing single crystal thin films be extended to other substances, including alloys of predetermined composition?*

9. REFERENCES

9.1. Reviews and General Articles

- Amelinckx, S., 1959, in *Methods of Experimental Physics*, Marton, L., editor (New York, Academic Press), Vol. 6, Part A, pp. 321-356.
- , and Dekeyser, W., 1959, in *Solid State Physics* (New York, Academic Press), Vol. 8, pp. 325-499.
- Barrett, C. S., 1952, 2nd edition, *Structure of Metals* (New York, McGraw-Hill), p. 92 ff.
- Bassett, G. A., Menter, J. W., and Pashley, D. W., 1959, in *Structure and Properties of Thin Films*, Neugebauer, C. A., et al., editors (New York, John Wiley and Sons), pp. 11-42.
- Brenner, S., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), pp. 157-188.
- Buckley, H. E., 1951, *Crystal Growth* (New York, John Wiley and Sons).
- Bunn, C. W., 1953, in *Fibers from Synthetic Polymers*, Hill, R., editor (Amsterdam, Elsevier Publishing Co.).
- Burke, J. E., and Turnbull, D., 1952, in *Progress in Metal Physics* (New York, Pergamon Press), Vol. 3, p. 220.
- Campbell, I. E., 1956, editor, *High Temperature Technology* (New York, John Wiley and Sons).
- "Crystal Growth," 1949, Discussions of the Faraday Society (London), Vol. 5.
- Doremus, R. H., Roberts, B. W., and Turnbull, D., 1958, editors, *Growth and Perfection of Crystals* (New York, John Wiley and Sons).

- Evans, R. C., 1939, *An Introduction to Crystal Chemistry* (Cambridge, Cambridge University Press).
- Fisher, J. C., Johnston, W. G., Thomson, R., and Vreeland, T., 1956, editors, *Dislocations and Mechanical Properties of Crystals* (New York, John Wiley and Sons).
- Flory, P. J., 1953, *Principles of Polymer Chemistry* (Ithaca, N.Y.; Cornell University Press), p. 563 ff.
- Fuller, C. S., 1959, in *Semiconductors*, Hannay, N. B., editor (New York, Reinhold), pp. 192-221.
- Hannay, N. B., 1959, editor, *Semiconductors* (New York, Reinhold).
- Hansen, M., 1958, *Constitution of Binary Alloys* (New York, McGraw-Hill).
- Hirsch, P. B., 1959, *J. Inst. Metals*, 87, 406.
- Holden, A. N., 1950, *Trans. Amer. Soc. Metals*, 42, 819.
- Inghram, M. G., 1959, in *Methods of Experimental Physics*, Marton, L., editor (New York, Academic Press), Vol. 6, Part A, pp. 29-38.
- Jamieson, J. C., and Lawson, A. W., 1959, in *Methods in Experimental Physics*, Marton, L., editor (New York, Academic Press), Vol. 6, Part A, pp. 407-437.
- Keller, A., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), pp. 499-532.
- Kittel, C., 1956, 2nd edition, *Introduction to Solid State Physics* (New York, John Wiley and Sons), Ch. 3.
- Kröger, F. A., and Vink, H. J., 1956, in *Solid State Physics* (New York, Academic Press), Vol. 3, pp. 307-435.
- Laves, F., 1955, in *Theory of Alloy Phases* (Cleveland, Ohio, American Society for Metals), p. 148.
- Lawson, W. D., and Nielsen, S., 1958, *Preparation of Single Crystals* (London, Butterworth's Scientific Publications).
- Mandelkern, L., 1956, *Chem. Reviews*, 56, 903.
- Mandelkern, L., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), pp. 467-495.
- Menter, J. W., 1958, *Adv. in Phys.*, 7, 299.
- , and Pashley, D. W., 1959, in *Structure and Properties of Thin Films*, Neugebauer, C. A., et al., editors (New York, John Wiley and Sons), pp. 111-147.
- Moffitt, W., and Ballhausen, C. J., 1956, *Ann. Rev. Phys. Chem.*, 7, 107.
- Nabarro, F. R. N., and Jackson, P. J., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), pp. 11-102.
- Neugebauer, C. A., Newkirk, J. B., and Vermilyea, D. A., 1959, editors, *Structure and Properties of Thin Films* (New York, John Wiley and Sons).
- Pashley, D. W., 1956, *Adv. in Phys.*, 5, 173.
- Pauling, L., 1955, in *Theory of Alloy Phases* (Cleveland, Ohio, American Society for Metals), p. 240.
- Pauling, L., 1960, 3rd edition, *The Nature of the Chemical Bond* (Ithaca, N.Y., Cornell University Press).
- Pfann, W. G., 1957, in *Solid State Physics* (New York, Academic Press), Vol. 4, pp. 424-521.
- Proceedings of the Symposium on High Temperatures*, 1956, (Menlo Park, California; Stanford Research Institute).

- Proceedings of an International Symposium on High Temperatures, 1958*, (Menlo Park, California, Stanford Research Institute).
- Shockley, W., Hollomon, J. H., Maurer, R., and Seitz, F., 1952, editors, *Imperfections in Nearly Perfect Crystals* (New York, John Wiley and Sons).
- Tanenbaum, M., 1959a, in *Semiconductors*, Hannay, N. B., editor (New York, Reinhold), pp. 87-114.
- , 1959b, in *Methods of Experimental Physics*, Marton, L., editor (New York, Academic Press), Vol. 6, Part A, pp. 86-121.
- Thurmond, C. D., 1959a, in *Semiconductors*, Hannay, N. B., editor (New York, Reinhold), pp. 145-191.
- , 1959b, in *Methods of Experimental Physics*, Marton, L., editor (New York, Academic Press), Vol. 6, Part A, pp. 39-85.
- Wells, A. F., 1950, 2d edition, *Structural Inorganic Chemistry* (Oxford, Clarendon Press).
- Wolfe, H. C., 1941, editor, *Temperature, Its Measurement and Control in Science and Industry* (New York, Reinhold), Vol. 1.
- Wolfe, H. C., 1955, editor, *Temperature, Its Measurement and Control in Science and Industry* (New York, Reinhold), Vol. 2.

9.2. Original Articles

- Ainslie, N. G., Hoffman, R. E., and Seybolt, A. U., 1960, *Acta Met.*, in the press.
- , Phillips, V. A., and Turnbull, D., 1960, *Acta Met.*, in press.
- Amelinckx, S., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), p. 139.
- Andrade, E. N., 1935, *Trans. Faraday Soc.*, **31**, 1137.
- Assmus, F., Detert, K., and Ibe, G., 1957, *Z. Metallkunde*, **48**, 344.
- Aust, K. T., and Rutter, J. W., 1960, *Trans. Met. Soc. AIME*, **218**, 682.
- Barber, D. J., Harvey, K. B., and Mitchell, J. W., 1957, *Phil. Mag.*, **2**, 704.
- Bassett, G. A., 1958, *Phil. Mag.*, **3**, 1042.
- , Menter, J. W., and Pashley, D. W., 1958, *Proc. Roy. Soc. (London)*, **A246**, 345.
- Beck, P. A., Holzworth, M. L., and Sperry, P. R., 1949, *Trans. AIME*, **180**, 163.
- Berghezan, A., unpublished.
- Bethe, H. A., 1929, *Annalen der Physik*, **3**, 133.
- Bolling, G. F., and Winegard, W. C., 1958, *Acta Met.*, **6**, 283.
- Bollmann, W., 1959, *J. Inst. Metals*, **87**, 439.
- Bradley, D. E., 1959, *Brit. J. Appl. Phys.*, **10**, 198.
- Burton, W. K., Cabrera, N., and Frank, F. C., 1951, *Trans. Roy. Soc. (London)*, **A243**, 299.
- Cabrera, N., 1957, in *Semiconductor Surface Physics*, Kingston, R. H., editor (Philadelphia, University of Pennsylvania Press), p. 327.
- , and Levine, M., 1956, *Phil. Mag.*, **1**, 450.
- , and Verziilyea, D. A., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), p. 398.
- Cahn, J. W., to be published.
- , and Hilliard, J. E., 1958, *J. Chem. Phys.*, **28**, 258.
- Cahn, R. W., 1950, *Proc. Phys. Soc. (London)*, **63A**, 323.
- Cech, R. E., 1956, *Trans. AIME*, **206**, 585.

- Chalmers, B., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), p. 291.
- Dash, W. C., 1957, in *Dislocations and Mechanical Properties of Crystals*, Fisher, J. C., et al., editors (New York, John Wiley and Sons), p. 57.
- , 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), p. 361.
- , 1959, *J. Appl. Phys.*, **30**, 459.
- Decker, B. F., and Harker, D., 1951, *J. Appl. Phys.*, **22**, 900.
- , and Kasper, J. S., 1959, *Acta Cryst.*, **12**, 503.
- Dunitz, J. D., and Orgel, L. E., 1957a, *J. Phys. Chem. Solids*, **8**, 20; 1957b, *ibid.*, **8**, 318.
- Dunn, C. G., and Walter, J. L., *Acta Met.*, to be published.
- , unpublished.
- Duwez, P., Willens, R. H., and Klement, W., 1960, *J. Appl. Phys.*, **31**, 1136.
- Ellis, W. G., Gibbons, D. F., and Treuting, R. G., 1958, in *Growth and Perfection of Crystals*, Doremus, R. H., et al., editors (New York, John Wiley and Sons), p. 102.
- Forty, A. J., 1957, *Proc. Roy. Soc. (London)*, **A242**, 392.
- Frank, F. C., 1949, *Discussions Faraday Soc.*, **5**, 49.
- , 1958a, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 3; 1958b, *ibid.*, p. 411; 1958c, *ibid.*, p. 304.
- , and van der Merwe, J. H., 1949, *Proc. Roy. Soc. (London)*, **A198**, 205.
- , and Kasper, J. S., 1958, *Acta Cryst.*, **11**, 184.
- Fullman, R. L., 1952, in *Metal Interfaces* (Cleveland, Ohio, American Society for Metals), p. 179.
- Geller, S., 1960, *J. Appl. Phys.*, **31**, 30S.
- Gilman, J. J., 1956, *J. Appl. Phys.*, **27**, 1262.
- Glossop, A. B., and Pashley, D. C., 1959, *Proc. Roy. Soc. (London)*, **A250**, 132.
- Goodenough, J. B., and Loeb, A. L., 1955, *Phys. Rev.*, **98**, 391.
- Gomer, R. J., 1958, *J. Chem. Phys.*, **28**, 458.
- Graham, C. D., and Cahn, R. W., 1956, *Trans. AIME*, **206**, 517.
- Green, R. E., Liebmann, B. G., and Yoshida, H., Brown University, unpublished.
- Guinier, A., and Tennevin, J., 1949, *Acta Cryst.*, **2**, 133.
- Ham, F. S., 1958, *J. Phys. Chem. Solids*, **6**, 335.
- Hart, E. W., 1959, *Phys. Rev.*, **113**, 412.
- Herring, C., 1951, *Phys. Rev.*, **82**, 87.
- Hillig, W. B., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 350.
- Hirsch, P. B., Kelley, A. J., and Menter, J. W., 1955, *Proc. Phys. Soc. (London)*, **68B**, 1132.
- , and Silcox, J., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 262.
- , Horne, R. W., and Whelan, M. J., 1956, *Phil. Mag.*, **1**, 677.
- Hoard, J. L., Hughes, R. E., and Sands, D. E., 1958, *J. Am. Chem. Soc.*, **80**, 4507.
- Hunter, L. P., 1958, Pittsburgh Diffraction Conference, unpublished.
- Iida, S., 1958, *J. Phys. Soc. (Japan)*, **12**, 222.
- Ives, M. J., 1958, Thesis, University of Bristol.
- Jackson, K. A., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 319.
- Jones, H., 1934, *Proc. Roy. Soc.*, **A144**, 225.

- Kelly, A., and Nutting, J., 1959, *J. Inst. Metals*, **87**, 385.
Lang, A. R., 1958, *J. Appl. Phys.*, **29**, 597.
———, 1959, *J. Appl. Phys.*, **30**, 1748.
Liebmann, B., Lücke, K., and Masing, G., 1956, *Z. Metallkunde*, **47**, 57.
Lighthill, M. J., and Whitham, G. B., 1955a, *Proc. Roy. Soc. (London)*, **A229**, 281; 1955b, *ibid.*, 317.
Lipscomb, W. N., 1959, *Advances in Inorganic Chemistry and Radiochemistry*, **2**, 117.
Loeb, A. L., 1958, *Acta Cryst.*, **11**, 469.
Lovell, L. C., Vogel, F. L., and Wernick, J. H., 1959, *Metal Progress*, **75**, 96.
Lücke, K., and Detert, K., 1957, *Acta Met.*, **5**, 628.
May, J. E., Walter, J. L., and Dunn, C. G., private communication.
Menter, J. W., 1956, *Proc. Roy. Soc. (London)*, **A236**, 119.
Mitchell, J. W., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 386.
Mott, N. F., 1948, *Proc. Phys. Soc. (London)*, **60**, 391.
Müller, E. W., 1958, *Acta Met.*, **6**, 620.
McClure, D. S., 1957, *J. Phys. Chem. Solids*, **8**, 311.
Newkirk, J. B., 1955, *Acta Met.*, **3**, 121.
———, 1958, *J. Appl. Phys.*, **29**, 995.
Nicholson, R. B., and Nutting, J., 1958, *Phil. Mag.*, **3**, 531.
Pashley, D. W., 1959, *Phil. Mag.*, **4**, 316.
———, and Presland, A. E. B., 1959, in *Structure and Properties of Thin Films* (New York, John Wiley and Sons), p. 199.
Pauling, L., Corey, R. B., and Branson, H. R., 1951, *Proc. Nat. Acad. Sci.*, **27**, 205.
Pearson, G. L., Read, W. T., and Feldman, W. L., 1957, *Acta Met.*, **5**, 181.
Price, P. B., Vermilyea, D. A., and Webb, M. B., 1958, *Acta Met.*, **6**, 524.
Rundle, R. E., 1957, *J. Phys. Chem.*, **61**, 45.
Schlier, R. E., and Farnsworth, H. E., 1958, *J. Phys. Chem. Solids*, **6**, 271.
Sears, G. W., 1953, *Acta Met.*, **1**, 457.
Smith, C. S., 1952, in *Metal Interfaces* (Cleveland, Ohio, American Society for Metals), p. 52.
Stein, R. S., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 549.
Turnbull, D., 1951, *Trans. AIME.*, **191**, 661.
———, 1957, *Acta Met.*, **5**, 502.
———, and Vonnegut, B., 1952, *Ind. and Eng. Chem.*, **44**, 1292.
Vermilyea, D. A., 1956, *J. Chem. Phys.*, **25**, 1254.
Walter, J. L., and Dunn, C. G., 1960, *Trans. AIME*, in the press.
Washburn, J., and Parker, E. R., 1952, *J. Metals*, **4**, 1076.
Watkins, G. D., 1959, *Phys. Rev.*, **113**, 79.
Webb, W. W., unpublished.
———, 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 230.
Wells, A. F., 1958, in *Solid State Physics* (New York, Academic Press), Vol. **7**, pp. 426-499.
Whelan, M. J., 1959, *J. Inst. Metals*, **87**, 392.
Wones, D. R., and Eugster, H. P., Geophysical Laboratory, Washington, D.C., unpublished.
Zener, C., 1948, *Trans. AIME*, **175**, 15.

Part VIII

MECHANICAL BEHAVIOR OF CRYSTALLINE SOLIDS

Chairman

Prof. EARL R. PARKER
*University of California
Berkeley, Calif.*

Members

Prof. A. H. COTTRELL
*University of Cambridge
Cambridge, England*

Prof. JOHN J. GILMAN
*Brown University
Providence, R. I.*

Dr. T. L. JOHNSTON
*Scientific Laboratory
Ford Motor Co.,
Dearborn, Mich.*

Dr. ARTHUR S. NOWICK
*International Business
Machines Corporation
Yorktown Heights, N. Y.*

Prof. J. WASHBURN
*University of California
Berkeley, Calif.*

CONTENTS

	Page
1. INTRODUCTION	456
2. DISLOCATION THEORY—ITS CURRENT STATE OF DEVELOPMENT AND LIMITATIONS.....	458
2.1. Foundations of the Theory.....	458
2.2. The Fine Structure of a Dislocation.....	460
2.3. Theory of Plastic Flow and Fracture.....	462
3. EXPERIMENTAL OBSERVATIONS ON FLOW AND FRACTURE	469
3.1. Flow Phenomena	469
3.1.1. Introduction	469
3.1.2. Work Hardening	472
3.1.3. Imperfections in Cold-Worked Materials.....	473
3.1.4. Flow at High Strain Rates.....	474
3.1.5. Effects of Grain Boundaries on Slip Distribution.....	475

	Page
3.1.6. Grain Boundary Sliding and Intergranular Cracking	476
3.1.7. Crack Formation at High Temperatures	477
3.1.8. Twinning	478
3.2. Fracture Phenomena	479
3.2.1. Introduction	479
3.2.2. Cleavage Fractures	480
3.2.3. Effect of Grain Boundaries on Cleavage	482
3.2.4. Ductile-Shear Fracture	482
3.3. Flow and Fracture of Two-Phase Systems	483
4. ANELASTICITY AND INTERNAL FRICTION	484
4.1. Introduction	484
4.2. Effects Caused by Point Defects and Solute Atoms	485
4.3. Effects Originating at Interfaces	487
4.4. Effects Associated with Dislocations and with Interactions Between Point Defects and Dislocations	488
4.5. Effects Associated with Phase Transformations	492
5. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH	492
5.1. Theory	492
5.2. Experiment	498
6. REFERENCES	494
6.1. Reviews and General Articles	494
6.2. Original Articles	495

MECHANICAL BEHAVIOR OF CRYSTALLINE SOLIDS

1. INTRODUCTION

The first real progress toward an understanding of the physical processes responsible for the plastic deformation and fracture of crystalline solids was made in the 1920's and 1930's. During this period, systematic studies on the crystallography of deformation and fracture were carried out on a variety of different metals and ionic solids. Much of this early work has been summarized by Schmid and Boas (1936R). A great advance in our understanding of the mechanical behavior of crystals came in 1934 when Taylor, Orowan and Polanyi independently suggested that the movement of dislocations was responsible for slip.

For nearly two decades thereafter, the theory of dislocations developed rather slowly. Not only were experimental techniques lacking for the direct observation of dislocations in crystals, but the further development of the theory was hampered by its great flexibility. It was usually impossible, for example, to choose among several possible dislocation models, each of which was theoretic-

cally capable of explaining a given set of experimental observations. Nevertheless, by 1953, a sufficient amount of information concerning the complex properties of dislocations in crystals had been accumulated to warrant the publication of two extended monographs, one by Cottrell (1953R) and the other by Read (1953R).

The period since about 1950 has been one of intensive and rapid advancement. This has been due in large measure to the introduction and development of new techniques for the observation of dislocations in crystals. Mainly through the use of dislocation etch pit and decoration techniques, and electron transmission microscopy, it has since become possible to study the actual arrangement of dislocations in annealed and cold-worked crystals, and even in complex commercial alloys, to observe the behavior of individual dislocations, and to study not only the interactions which occur between dislocations themselves but also interactions involving dislocations and other structural defects, such as impurity atoms, vacancies, grain boundaries, etc. As a result, most of the basic postulates of the theory have now been confirmed and many of the dislocation reactions predicted by theory have been verified, so that today the theory rests on a firm experimental foundation.

With the continued refinement of the techniques presently available and with systematic application of these techniques, it should now be possible to test the correctness or uniqueness of a given dislocation model and to correlate changes in mechanical properties with changes in structure on a much finer scale than has been achieved heretofore. Although much has already been accomplished in this direction, the task which still remains is enormous and many important subjects or areas of interest are as yet virtually unexplored. In this report, an attempt has been made to summarize and evaluate our current understanding with regard to factors which control the mechanical behavior of crystalline solids.

Following this brief introduction, we present in section 2, a discussion of the current state of development and limitations of dislocation theory. Experimental observations on plastic flow and fracture are reviewed and discussed in section 3 and this is followed by an evaluation of our knowledge of anelastic effects and internal friction phenomena in metal and alloy systems. In each section, the existing information is reviewed, the important problems requiring further attention are emphasized and what appear to be promising paths for future research are indicated. A brief

summary which incorporates our major conclusions and suggestions for future research will be found in section 5.

2. DISLOCATION THEORY—ITS CURRENT STATE OF DEVELOPMENT AND LIMITATIONS

2.1. Foundations of the Theory

The theory of dislocations began as a topic in elasticity, where it was concerned with the states of self-stress associated with multivalued displacements in multiply-connected elastic solids. This article, however, is concerned with the theory only in its modern sense, in which a dislocation is regarded as a certain kind of defect in a crystal and is definable by means of a Burgers circuit. The discussion is further confined to the influence of such dislocations on the mechanical properties of crystalline solids.

Dislocations are geometrically complex defects, but the theoretical foundations of dislocation theory (Cottrell, 1953R; Nabarro, 1952R; Read, 1953R; Friedel, 1956R; Seeger, 1955R, 1958R; Kröner, 1958R) are simple and not subject to real doubt. From the theoretical point of view, dislocations are natural consequences of the periodicity of the crystalline field of interatomic forces that allow mechanically stable patterns of atoms or molecules to exist in which, for example, $n + 1$ particles in one layer are matched against n in the next. The justification for this is based on ordinary classical mechanics. The arguments for the existence of dislocations and for some of the general features or properties of dislocations do not depend upon the physical nature of the interatomic forces. For this reason, a raft of soap bubbles or a lattice of magnets makes a valid model for demonstrating dislocations.

To verify that real crystals can be dislocated, it is sufficient to observe that a ductile crystal can be bent plastically so that it is made to slip in opposite directions at opposite ends of its slip planes. A Burgers circuit taken around the outside of this crystal would then show that there were dislocations in the crystal and would give a value for their total Burgers vector.

Strictly, however, this proves only that a "state of dislocation" exists within the crystal rather than distinct, individual, dislocations. The theoretical justification for the latter also depends on the periodicity of the crystalline field, for this not only stabilizes the gradient of plastic shear along a bent slip plane, but it also breaks up this gradient into a sequence of unit steps each of which corresponds to a separate dislocation line in the crystal. Thus, apart from "cavity dislocations" which formally represent cracks

in elastic solids (Cottrell, 1959), crystals do not contain continuous distributions of infinitesimal dislocations (although it is often mathematically convenient to pretend that they do) but finite collections of discrete dislocations. In the early days of the theory, this was taken for granted, but it is, in fact, not really self-evident. There may be materials—perhaps certain types of liquid crystals—in which infinitesimal dislocations are continuously distributed. The state of dislocation in a sliding boundary is usually regarded as continuous. In a crystal, however, as Peierls (1940) first proved, the cohesive forces across a slip plane normally prevent the center of a dislocation in that plane from spreading out beyond a few atomic spacings. This then is the theoretical justification for the view that discrete dislocations exist in crystals.

Recent experimental work on the direct observation of dislocations in crystals has fully confirmed these basic ideas of the theory. In fact, it is now no longer necessary to introduce the fundamental concepts of the theory from the above point of view, strong though the justification for this approach. Instead, the many ingenious techniques which have been developed and applied to the study of small-angle dislocations boundaries, dislocation etch pits, the decoration of dislocations by foreign atoms and precipitates, electron-microscopy and X-ray scattering, make it possible to develop the subject from a completely empirical, observational standpoint. It is now possible to see edge and screw dislocations as well as dislocation nodes and networks; to see dislocations glide, pile up, and curve around obstacles; to see the cross-slip of screw dislocations and the climb of edge dislocations; to see the segregation and precipitation of foreign atoms on dislocations; and to observe the creation of dislocation rings from clusters of lattice vacancies (Hirsch, 1959R; Amelinckx and Dekeyser, 1959R). Frank-Read sources have been photographed (Dash, 1957), and the stresses to move dislocations through crystals at known velocities are beginning to be measured by direct observation (Johnston and Gilman, 1959; Stein and Low, 1960). The dissociation of dislocations into partials, separated by stacking faults, has been seen and studied, particularly in stainless steel and in some copper alloys (Hirsch, 1959R). Studies of the elastic strain fields associated with dislocations and the elastic interactions between dislocations can be greatly facilitated by observations of the positions of dislocations at small-angle boundaries, at the junctions of slip planes, and in piled-up groups.

Photoelastic techniques have been used recently to study stresses around individual dislocations in silicon (Bond and Andrus, 1956).

The rate of progress in both theory and experiment has been so rapid in recent years that seemingly serious obstacles to the further development of the theory may quickly disappear. Extensive portions of the subject have already been developed satisfactorily. For instance, there seems to be little more to do regarding the long-range properties of dislocations in crystals except to repeat, for more complicated crystal structures, the kinds of analyses that have already been made for simple crystals (Cottrell, 1953R).

2.2. The Fine Structure of a Dislocation

However, a major barrier to progress still exists with regard to the fine structure, i.e., the detailed arrangement of atoms, at the center or core of a dislocation. Until substantial progress has been made in this direction, we will never be absolutely certain that we understand, for example, why diamond is hard and brittle, or why aluminum is soft and ductile, or why face-centered cubic metals prefer to slip on $\{111\}$ planes even though their dislocations have lower elastic energies in $\{110\}$ planes; nor will we be able to decide or predict how fast a dislocation can move through a crystal under a given applied stress.

A fundamental limit to the velocity of a dislocation exists, of course, at the velocity of propagation of elastic waves in the material. A few years ago, an elegant dynamical theory of "relativistic" dislocations was developed in which the concepts of the mass, momentum, and kinetic energy of a dislocation were critically examined and defined (Frank, 1949; Eshelby, 1949). The theory promised to be important practically because a dislocation travelling almost at the velocity of sound would have enough momentum to push through many otherwise impenetrable obstacles placed in its path, and it would have enough kinetic energy to create new dislocations whenever favorable opportunities for shedding some of its energy presented themselves. In fact, it seemed that if such a dynamical slip process were once started, it would be most difficult to stop.

However, the interactions of atoms at the center of a moving dislocation, the scattering of phonons by the dislocation, and the radiation of elastic energy from supersonic, short-wave Fourier components of the elastic wave-packet which represents the dislocation, all provide a resistance to motion that becomes quite appreciable as the dislocation velocity approaches that of sound. These effects tend to limit the velocity of the dislocation to the

subrelativistic range. In spite of determined theoretical attacks, questions concerning the behavior of fast-moving dislocations have not yet been clarified, mainly because the approximations used in the analytical treatments of the problem begin to break down seriously when the dislocation velocity reaches the interesting range of speeds. As a result, practically all recent theories of crystal plasticity have assumed that relativistic effects can be disregarded (Liebfried and Haasen, 1954; Seeger, 1954), not because any firm proof of this exists, but simply because it has not been found absolutely necessary to appeal to such effects. This represents a very weak point in the subject and, even though it is beginning to look now as if the "quasistatic" approach will eventually prove to be the correct one for most of the important problems, a determined effort to clear up this phase of theory would be welcomed.

Before such a problem could be tackled at the level that now seems to be required, it would be necessary first to develop improved theoretical treatments of the Peierls-Nabarro force, of the splitting of dislocations into partials, of the conditions for cross-slip, of the force required to move a jogged dislocation, and of dislocation damping phenomena (Seeger, 1957). So far, only in close-packed crystal structures, where splitting into partial dislocations in close-packed planes is a particularly obvious process, has much real progress been made, and even here many problems remain. For example, a great deal of evidence suggests that stacking faults have relatively high energies in aluminum and low energies in copper (Fullman, 1951; Seeger and Schoeck, 1953; Schoeck and Seeger, 1955), yet dislocation loops formed by the condensation of quenched-in vacancies in copper often resemble those in aluminum rather than those expected in a metal of low stacking fault energy (Hirsch, 1959R). Again, Bordoni damping in plastically strained copper remains full of perplexities (Bordoni, 1954; Bordoni et al., 1959; Lücke and Granato, 1957). With regard to the traditional toughness of f.c.c. metals, no one has yet shown whether a metal such as copper could sustain a brittle cleavage crack propagating at maximum speed on the most favored plane, or whether the crack front would transform into a shower of glide dislocations. For the b.c.c. metals and for nonmetals such as diamond or ionic crystals, the theory of the fine structure of dislocations dwindles away into a miscellany of intuitive ideas based partly on notions about directed bonds and partly on the fact that the yield strengths of such materials are highly sensitive to the temperature and the rate of straining.

The central difficulty connected with further development of the theory lies beyond the special nature of dislocations themselves. It is part of the general many-body problem in which particle-particle interactions cannot be ignored by the use of a continuum approximation and, likewise, nonlinear terms in those interactions cannot be disregarded. The same problem occurs in the development of theories of radiation damage, sputtering, chemical kinetics, and the viscosity of liquids. However, the great advances in computer techniques which have been made in recent years should soon make it possible to solve problems of this kind. For dislocation problems, there would seem to be two stages to such computations. First, to program computers to make atom-by-atom calculations of the structures of dislocation cores in those materials for which reasonably reliable force laws are available. Fortunately, it should be possible to include copper among these. Second, to develop the equivalent laws of interatomic force for other materials such as the body-centered cubic metals. This will be a difficult task because these forces cannot be of the simple, central, nearest-neighbor type. A useful approach might be to use the force constants derived from X-ray diffuse scattering studies of the lattice vibrations. Until the structures of dislocations have been computed using such realistic force laws, however, we shall never be able to develop a really complete theory of the mechanical properties of materials such as iron.

There are several opportunities for experimental work on the centers of dislocations. Measurements of the velocity of dislocations, as a function of stress and temperature, in good crystals are particularly important. Only limited data are as yet available on only two materials, LiF and silicon-iron (Johnston and Gilman, 1959; Stein and Low, 1960). Accurate data concerning the flow stresses of high-purity materials, as functions of temperature and strain rate, are also important, as is detailed information about the nature of Bordoni damping. Careful studies of the electrical and chemical properties of dislocations, particularly in nonmetallic crystals, should also prove informative.

2.3. Theory of Plastic Flow and Fracture

By far the greatest part of dislocation theory is concerned with understanding the mechanical properties of solids. The work which has already been carried out and which still remains to be done in this field is immense because of the large number and variety of processes that occur. The various types of flow, hardening, and fracture processes need to be analyzed for all the various

types of crystal structures and types of solids. These processes involve the interactions of dislocations with one another, at both long and short range, in parallel and intersecting glide systems, and in networks and boundaries of various types; interactions with solute atoms, substitutional and interstitial, dilute and concentrated, mobile and fixed, random, ordered, and segregated or clustered; interactions with precipitates, coherent and noncoherent, fine and coarse, platelike and spheroidal; interactions with point defects such as vacancies and clusters of point defects; interactions with grain boundaries and with the material on the other side of such boundaries.

The difficulties in the field are, in fact, due mainly to this profusion. When faced with some experimental observation, e.g., the formation of extrusions and intrusions on the surface of a fatigued metal (Thompson, 1959; Wood, 1959), or serrations on a tensile stress-strain curve (Cottrell, 1953R), the problem is usually not to devise possible explanations of the effect but to prove which, if any, of these explanations is correct. Fortunately, the techniques now available for observing dislocations are highly suitable for this kind of work and are already playing a decisive part in establishing the actual processes of deformation and fracture in materials. Furthermore, they have the great advantage of enabling the individual events of plastic flow to be seen directly, so that it is no longer necessary to work with specially purified and simple materials in order to avoid ambiguities of interpretation. Polycrystalline stainless steel is as good a subject for such studies as a single crystal of super-purity aluminum. The principles of creep resistance can now be sought by direct observation of complex creep-resistant alloys, and the long upward journey through artificial, highly simplified models or versions of such materials may now no longer be necessary.

The extent to which the Frank-Read mechanism of dislocation multiplication actually operates in crystals has recently been questioned. Certainly, there is no doubt that it sometimes functions, because photographs of Frank-Read sources in operation, e.g., in silicon, have been taken (Dash, 1957). However, transmission electron microscopy has shown that grain boundaries often act as sources of dislocations in thin-film specimens (Hirsch, Horne, and Whelan, 1957), and studies on alkali halide crystals, particularly LiF, have shown that dislocations usually multiply by a process of double cross-slip (Gilman and Johnston, 1957; Johnston and Gilman, 1960).

An important recent advance in the theory of work hardening

has been the realization, supported by electron microscopical observations (Thornton and Hirsch, 1958; Hirsch, 1959R; Saada, 1960) that "forest" dislocations can, by combining with glide dislocations, give rise to a strong thermally-insensitive component of work hardening; at a given temperature, this contribution to work hardening is proportional to and superimposed upon the thermally-sensitive component already attributed to the "forest" (Cottrell and Stokes, 1955). An explanation is thus provided for the well-established law of proportionality between these two contributions to work hardening. The observations also help to clarify the reason for the irreversibility of plastic deformation, and provide a mechanism for work hardening even in materials in which sessile partial dislocations are not expected to form. On the other hand, the work softening which occurs when a cold-worked metal is strained again at a higher temperature (Cottrell and Stokes, 1955) provides incontestable evidence that the states of cold work (distributions of dislocations) developed at different temperatures differ not only in intensity but also in kind; this strongly suggests that no single parameter, such as the density of intersecting dislocations, could by itself adequately define a state of cold work. It would appear that piled-up groups of dislocations and breakouts from these groups cannot be disregarded. However, if a measurable part of work hardening is attributed to the heterogeneous grouping of dislocations into pileups, the gratifying manner in which the pure forest theory can so easily account for the proportionality of the thermally-sensitive and thermally-insensitive components is then lost. The problem now confronting the theory is to find a convincing means of reconciling these two sets of experimental facts.

Alternating plastic deformation produces many perplexing effects. Now that the screw-screw mechanism for generating lattice vacancies has been opened to doubt (Cottrell, 1957a), it is no longer clear how point defects are actually produced by gliding dislocations. On the other hand, experimental evidence, particularly from fatigue experiments, leaves little doubt that point defects are produced as a result of plastic deformation. The formation, internal structure, and properties of heavy slip bands in fatigue is another problem that requires further study, especially since fatigue cracks usually start from places where such bands meet the surface of the specimen. The mechanism by which extrusions and intrusions are formed may be illuminated by metallographic studies of taper sections taken from the surface layers of fatigued specimens (Wood, 1959).

Good progress has been made and is currently being made in the study of alloy hardening. Random solid-solution hardening seems to be broadly understood, although further work is needed to clarify the role played by the interactions of solute atoms with dislocations, dislocation networks, and grain boundaries. The very strong hardening produced by interstitial elements in body-centered cubic transition metals is better understood than is alloy hardening in substitutional face-centered cubic solid solutions. Recent experiments have indicated that hardening in substitutional solid solutions is not due to solute atoms randomly dispersed in the perfect part of the crystal, but rather to the foreign atoms segregated at dislocations, subgrain, and grain boundaries (Parker, 1958). Techniques have not yet been devised which will yield quantitative information about the concentration or energy state of atoms segregated at such crystal imperfections, although thin-film transmission electron microscopy appears to offer some promise in this regard. A quantitative theory of solution hardening is badly needed, but there is no indication that an advance of this kind will be forthcoming in the near future.

Measurements of dislocation velocities as a function of the applied stress and temperature would be useful. Many examples of the pinning of dislocations by segregated foreign atoms have now been demonstrated using the presence of a sharp or discontinuous yield point, as well as strain aging and internal friction phenomena, as indicators of the effect. Independent evidence for the segregation process can sometimes be obtained from etching experiments in which etch pits are formed on dislocations only when certain kinds of impurity are present in the material and when an aging treatment is given to allow these impurities to diffuse into dislocations. Theoretical considerations have shown that the yield point of a dislocation locked by impurities should vary sensitively with temperature (Cottrell, 1953R), and various pieces of evidence have been gained in support of this argument. It was thought originally that the temperature dependence of the yield strength of iron could be attributed entirely to dislocation locking by impurities, but more recent work has shown that this is not so, and that a substantial part of effect of temperature must be ascribed to a lattice friction which may be a Peierls-Nabarro force (Heslop and Petch, 1956; Petch, 1959). Further work is needed to establish firmly the nature and characteristics of this frictional force.

The principles of hardening in ordered alloys have been examined, and some interesting facts have emerged. It has been

shown, from both theory and experiment, that the hardening is greatest when the alloy is in a partly ordered state, this structural condition being associated either with the presence of a certain degree of short-range order or with long-range order accompanied by a fine domain size. At higher temperatures, where the solute atoms diffuse more rapidly, a form of hardening occurs which has been ascribed to the ordering of solute atoms near the center of a moving dislocation. This type of hardening is rather similar to the "blue brittleness" and "Portevin-LeChatelier" effects, both of which are believed to result from the segregation of mobile solute atoms to moving dislocations (Cottrell, 1953R). Advantage may perhaps be taken of such effects in improving the creep resistance of certain materials.

Precipitates and their interactions with dislocations are particularly suitable subjects for study with the electron microscope. Replication methods as well as thin-film transmission techniques have been used to elucidate the detailed nature of the interactions which occur between dispersed phases and dislocations. As a result of such studies, and also of careful analysis of the temperature dependence of the flow stress in precipitation-hardened alloys, considerable progress has been made recently. It has been shown that dislocations glide through coherent precipitates, but not through incoherent particles, under the influence of an applied stress and thermal activation (Nicholson, Thomas and Nutting, 1959). Also, the separation of mutually attracting atoms, which occurs when coherent precipitates are cut by dislocations, leads to an important source of hardening which is independent of the elastic hardening due to the strain field around the precipitate. Sometimes the strain field surrounding a large particle is so intense that a cluster of dislocation loops is created around the particle to accommodate the strain. The range of the particle as an obstacle to slip is then effectively increased, especially when these dislocation loops attract to themselves other precipitates or solute atoms which anchor them more firmly in position.

The importance of anchoring dislocations by means of suitably distributed precipitates is now generally recognized as a guiding principle in the design of creep-resistant alloys. Creep resistance is lost at high temperatures when dislocations held up by obstacles in slip bands become capable of climbing rapidly from one slip plane to another, where they may meet and annihilate dislocations of opposite sign. This process can be delayed or prevented by anchoring the dislocations with immobile segregates or precipitates. It is expected that future work in this field will be

directed towards learning how to produce thermodynamically and structurally stable precipitates in alloys and how to distribute them where they can be most effective.

In recent years, considerable interest has been shown in the hardening of metals through irradiation damage. The effects can be very strong, and the yield strengths of metals such as copper, aluminum, and steel have been raised by irradiation up to the highest level that can be achieved with such materials in the precipitation-hardened state (Blewitt and Coltman, 1951; Kunz and Holden, 1954). Experiments have shown that two distinct processes of radiation hardening exist. The effect of small doses of irradiation in suppressing dislocation damping and in changing the shape of the stress-strain curve strongly suggests that some pinning of dislocations by segregated point defects takes place. The mechanism of pinning is not yet clear; it may, for instance, be due to clustered defects along the dislocation line or to jogs on the dislocation produced by the absorption of point defects generated during irradiation. Electron microscopy of quenched metals and alloys has shown, recently, how strongly the shapes of dislocation lines can be affected by quenched-in vacancies. Large jogs are formed on edge dislocations, and screw dislocations become wound up into spirals (Hirsch, 1959R).

Nevertheless, it seems that the main component of radiation hardening in severely hardened specimens is caused by dispersed point defects, presumably in the form of small clusters (Cottrell, 1957b). The evidence supporting this view is derived from several types of experiments: from the shape of the stress-strain curve; from the variation of yield stress with grain size, which has shown that, for iron at least, prolonged neutron irradiation raises the "frictional" stress in the slip planes by an amount that is independent of both the grain size and the temperature of straining (Hull and Mogford, 1958); and from experiments, mainly on lithium fluoride, in which the difficulty of moving dislocations through an irradiated crystal has been proved directly by means of etch pit techniques (Johnston and Gilman, 1959).

Although many points connected with radiation hardening still require clarification—for example, the effect of irradiation at very low temperatures, the work-softening that occurs in irradiation-hardened copper, or the influence which the passage of dislocations exerts upon point defects distributed in their slip planes—these items, nevertheless, appear to be matters of detail; and it seems doubtful whether the study of radiation hardening will reveal any major aspects of the behavior of dislocations that

could not be deduced from studies of alloy, quench and work hardening. Thus, the effect of radiation on the transition temperature of steel appears to be quite normal, and other forms of dispersion hardening such as quench-age hardening produce similar increases in the transition temperature for the same amount of hardening.

The role which dislocations play in processes of fracture is now attracting a good deal of interest (Averbach et al., 1959R), especially since recent experiments with ionic crystals have provided direct proof that cracks can be formed at the ends of slip planes where dislocations have run up against obstacles such as groups of dislocations on intersecting slip planes. There are good opportunities to extend such observations to other materials, especially to those metals that fail by cleavage at low temperatures. Development of the theory of the coalescence of glide dislocations and their transformation into "cavity dislocations" or cracks (Cottrell, 1959) is greatly impeded by our inability to handle precisely problems concerning the centers of dislocations. There can be little doubt that cracks must be formed when a number of edge dislocations run together to form a single edge of large Burgers vector; however, during the initial stage of this process, in which the two leading edge dislocations coalesce, it is not possible to conclude with certainty, on the basis of current knowledge, whether the concentration of strain energy produced by the coalescence will be dispersed by nearby slip or by fracture. The difference in fracture behavior between the b.c.c. and f.c.c. metals may well depend on this question.

The extent to which dislocations play a direct part in the nucleation or propagation of ductile fractures and fractures encountered under creep conditions is not yet clear. Some types of ductile fracture are nucleated in the form of plastic cavities, usually situated on foreign particles or inclusions, and grow by a process of localized plastic flow when the rate of work hardening becomes sufficiently low. Plastic instability seems to be the critical feature underlying processes of ductile fracture, and dislocations are important mainly insofar as their behavior leads to this instability. The role of dislocations in creep fractures is even more obscure because the primary processes in these fractures appear to be grain boundary sliding and vacancy diffusion in the vicinity of grain boundaries (Grant, 1959; Gifkins, 1959R). The plastic properties of the grains are, of course, important in connection with the relief of strains produced at triple points by grain boundary sliding. If the dislocations in the three contiguous

grains are too firmly locked, these strains cannot be accommodated except by fracture in the grain boundaries at the junction points between the grains.

Finally, it is useful to consider the relation between the theory of dislocations and the formal mathematical theory of plasticity, since a dislocation can be regarded as a unit of plastic strain in an elastic-plastic solid. Very few solutions of elastic-plastic strain fields have been obtained so far, using the classical methods, and it may be asked whether dislocation theory could help to provide more solutions. Some problems already solved in dislocation theory—for example, the distribution of dislocations in a bent or twisted specimen to avoid far-reaching elastic strains—correspond to unusual and difficult problems in the formal theory. Further development of this aspect of dislocation theory would be particularly useful if it led to reliable solutions—even though they might be only approximate ones—of otherwise intractable problems in the theory of elastic-plastic solids. An obvious application would be to the study of the growth of plastic zones around stressed notches.

3. EXPERIMENTAL OBSERVATIONS ON FLOW AND FRACTURE

3.1. Flow Phenomena

3.1.1. Introduction

The theories of plastic deformation and fracture proposed in the classic papers by Taylor (1934) and by Griffith (1921) still form the basis of modern concepts. It seems clear that plastic deformation takes place by the motion and multiplication of dislocations, that strain hardening is caused by the interactions between dislocations, and that the Griffith criterion in a modified form is a necessary condition for fracture. However, this general knowledge does not lead to a clear understanding of why such striking differences exist in the mechanical properties of various materials, or even why a given substance behaves so differently under various states of stress, or under various conditions of temperature, environment, and previous thermal or mechanical history.

The great advances in understanding that have been made during the last 25 years have been due in large part to the continuous refinement of the early ideas. These concepts have been developed in enough detail so that some significant correlations involving diverse phenomena have now emerged. However, the

interactions which occur between various kinds of point defects in crystals, and between individual point defects, clusters of point defects and dislocations, are often so complex (Hirsch, Horne and Whelan, 1957; Hirsch et al., 1958; Hirsch, 1959R) that satisfactory theoretical explanations for many observations have not yet been developed. It is now generally recognized that experiments must be conceived, performed, and interpreted with extreme care, because variables other than those anticipated often influence the results. Much of our present knowledge has come from experiments on high-purity metal crystals, but there is good reason to believe that if ionic materials could be obtained in a state of purity and structural perfection comparable with present-day silicon and germanium, future work on these materials would add materially to our understanding of flow and fracture phenomena.

The flow and fracture characteristics of any crystal are partially determined by events that occur during growth of the crystal, since the concentrations as well as the distribution of impurities, vacancies, and dislocations in the resulting crystal are affected by the growth conditions. Only with silicon has it now become possible to grow large, essentially perfect crystals which do not contain any dislocations and in which the impurity concentration is as low as one part in 10^9 (Dash, 1958). Of necessity, all experiments dealing with the deformation and fracture behavior of bulk metal and ionic crystals have so far been carried out on relatively imperfect and impure materials, and the interpretation of experimental results has almost always been beclouded by a lack of knowledge of the real purity or structural perfection of the specimens employed.

It is generally believed that crystals, particularly those grown from the melt, nearly always contain a three-dimensional network of dislocations which are introduced as a consequence of the stresses developed during growth (Washburn, 1958; see also Part VII, section 5.3). In practice it is almost impossible to avoid such stresses. They arise because of the presence of curved isothermal surfaces resulting from anisotropic heat loss to the surroundings. Mobile dislocations which exist in the newly formed crystal apparently multiply during growth because of the thermal stresses generated during cooling. At the high temperatures involved, the dislocations are able to form metastable networks which persist when the crystal is cooled to room temperature. The detailed structure of these networks has been studied in only a few materials, particularly ionic crystals, mainly

through the use of decoration techniques (Amelinckx, 1957, 1959R). The new technique of transmission electron microscopy (Hirsch, Horne and Whelan, 1956, 1957; Hirsch, 1959R) may eventually provide detailed information concerning the dislocation densities and network structures in a wider variety of materials. More importantly, this technique may help to establish correlations between the dislocation structure and the growth conditions or thermal-mechanical history. However, there is still no direct way to detect individual point defects or small clusters of point defects. Their distribution is certain to be sensitive to the growth conditions and thermal history, and probably has an important influence on the minimum stress required for dislocation motion.

It is often assumed that plastic deformation is initiated by the stress-induced motion of some elements of the three-dimensional dislocation network. However, the role of the grown-in-network of dislocations as the source of slip is now in doubt, because etch-pit techniques have provided evidence that, at least for some ionic crystals, grown-in dislocations are so strongly locked by impurities and noncrystallographic trajectories that they are never able to move. Prismatic dislocation loops formed by the collapse of disc-shaped vacancy clusters represent, in theory, another possible source of mobile dislocations (Kuhlmann-Wilsdorf, 1958; Fisher, 1957; Weertman, 1957b) but their importance has not yet been established in real materials. It seems unlikely, considering the diverse conditions employed during crystal growth and subsequent thermal treatment, that such loops could always be present. Specimen quality control has usually not been good enough to confidently rule out either external surface damage or minute inclusions of foreign particles as important sources of slip (Gilman, 1959; Johnston and Gilman, 1960). Nevertheless, whatever the true nature of the slip sources, it appears that they are sometimes not present in very small crystals. Micron-size crystals or whiskers of many metals and ionic materials have been obtained, for example, which will support shear stresses close to the theoretical strength without any evidence of plastic deformation (Brenner, 1958R). Critical experiments are needed to define more clearly the nature of dislocation sources in crystals.

The yield strength, as ordinarily measured by extrapolating the plastic portion of the stress-strain curve to zero strain, does not correspond to the stress at which dislocations first begin to move. Substantial amounts of plastic flow occur at stresses well below the "knee" of the stress-strain curve, and, indeed, con-

siderable uncertainty still exists regarding what actually happens in this lower stress region. The stress level at which the stress-strain curve bends over, i.e., the flow stress, varies with the density and distribution of dislocations, point defects, defect clusters, and with the impurity concentration. The hardening effect produced by a given concentration of impurity atoms also appears to depend on the nature of the dislocation substructure which is present. Neither impurity locking of dislocation sources nor the frictional resistance to the motion of dislocations that may result from short-range order appears to satisfactorily explain certain forms of solid solution hardening.

3.1.2. Work Hardening

In hexagonal close-packed crystals, and when plastic strain is confined to a single slip system in face-centered cubic crystals, plastic deformation causes relatively little hardening (Lücke and Lange, 1952; Seeger, 1957, 1958R). The small amount of strengthening that does occur is apparently due to the formation of positive-negative dislocation pairs. This kind of dislocation interaction does not cause strong hardening, and the structural changes accompanying deformation can be completely eliminated by heating the specimen to a temperature that will permit dislocation climb to occur (Edwards et al., 1953). When the strain distribution is not uniform and local lattice curvatures are produced during deformation, the crystal can never be returned to its initial state by annealing because of the metastable system of small-angle boundaries which forms in regions of high lattice curvature. Nonuniform deformation always occurs when polycrystalline specimens are deformed (Amelinckx and Dekeyser, 1959R), and this condition is very difficult to avoid even in careful experiments on single crystals. The complex dislocation networks that arise in polycrystalline materials cause the flow stress to increase markedly with increasing strain (Parker, 1958).

The strong work hardening characteristics of face-centered cubic crystals are clearly the result of glide on more than one slip plane. Segments of sessile dislocation of the Lomer-Cottrell type are thought to provide strong enough obstacles to allow the formation of large piled-up groups of dislocations, thereby creating high internal stresses (Cottrell, 1953R; Mott, 1960R). The maximum amount of strain hardening that can be obtained is apparently determined by thermally-activated cross-slip of the leading dislocations in the pile-ups, and is presumably a function of the stacking fault energy of the material (Seeger, 1958R).

Transmission electron microscopy has revealed, for example that in aluminum, which presumably has a high stacking fault energy, (Hirsch, Horne and Whelan, 1957; Hirsch, 1959R), plastic strain results in the formation of a system of relatively perfect subgrains separated by small-angle boundaries. The subgrain size is also found to decrease with increasing cold work. On the other hand, in face-centered cubic metals with low stacking fault energies and in body-centered cubic materials, an increasing fraction of the total volume appears to become filled with a complex tangle of dislocation lines as the amount of cold work is increased (Hirsch, 1959R).

Although the arrangements of dislocations in work-hardened crystals, as seen in the electron microscope, appear to be quite different in different materials, there is some uncertainty about why this should be so. The materials studied thus far have not been examined at equivalent temperatures, and until this has been done, general conclusions drawn from these observations may be subject to some criticism. The theory of strain hardening has not been developed satisfactorily for body-centered cubic materials. Existing theories of work hardening are, in fact, so highly idealized that they do not accurately represent real situations. What seems to be needed most are quantitative measurements of the stresses required to make slip penetrate intersecting glide planes and grain boundaries. When such fundamental information becomes available, then the formulation of a sound theory may become possible. At present, an experimental approach appears to offer most promise and transmission electron microscopy, in particular, may be of great aid in this regard.

3.1.3. Imperfections in Cold-Worked Materials

To understand the nature of the cold-worked state, detailed information is needed concerning the types of lattice defects that are present in such materials, how these defects are generated during deformation and in what concentrations, and how they can interact to affect the mechanical properties. Some properties are also extremely sensitive to small concentrations of specific impurities; hence, careful control of experiments is necessary in order to obtain meaningful results.

Vacancies appear to be generated in fairly large numbers by plastic flow. Boas (1957) has reported vacancy concentrations as high as one per cent in nickel deformed at room temperature. Although isolated vacancies have a significant effect upon the electrical resistivity, they seem to exert little influence upon the

hardness. The vacancies produced by cold work anneal out at relatively low temperatures, i.e., at about 200–300°C in nickel and below room temperature in copper, yet recrystallization does not occur in nickel until temperatures in the neighborhood of 600°C are reached. The fact that the recrystallization temperature is increased by the presence of impurity atoms is well known, but it is not as widely appreciated that the lower temperature recovery process, which is presumably associated with the disappearance of vacancies, is actually retarded by impurity atoms (Clarebrough, Hargreaves and West, 1956).

Boas (1957) has also found, by measuring the changes in internal energy which take place during the annealing of cold-worked metals, that the density of dislocations increases almost linearly with strain. The interpretation of stored energy data is, however, subject to many uncertainties, particularly because it is not possible to separate unambiguously the contributions to the internal energy arising from the various types of defects which are present in cold-worked material, or from the various possible relaxation processes involving these defects. The use of newly developed techniques in electron microscopy may be expected to clarify some of the problems involved in studies of imperfections generated by cold work.

Stacking faults are produced by cold working close-packed metals. Barrett (1952) and others have demonstrated by means of X-ray techniques that stacking faults are present in cold-worked metals, and Hirsch et al., (1957) has observed stacking faults in gold and other metals by transmission electron microscopy.

Nevertheless, few direct measurements of the concentrations and distributions of these defects have been made. Moreover, much still needs to be learned about the influence of particular types of lattice defects on the mechanical behavior. It is known in a general way that such imperfections cause large property changes, but the effects associated with specific imperfections are known in only a few instances.

3.1.4. Flow at High Strain Rates

This is a field which is now ready for intensive exploratory investigation.¹ In the past, most of the effort has been concerned with the development of experimental techniques and with the mechanics of plastic wave propagation. Flow stresses at various

¹ See, for example, the conference on Response of Materials to High Velocity Deformation, sponsored by A.I.M.E., July 11, 12, 1960, to be published by Interscience Publications, Inc., New York.

rates of strain are known for some materials, especially steels (Manjoine, 1944) and aluminum (Hauser, Simmons and Dorn, to be published), but very little information is available about more fundamental quantities. High-speed dislocation motions have been studied in only a few materials, e.g., lithium fluoride. Almost any systematic experiments in this field will be useful, and especially experiments in which individual dislocation motions are observed. Data on dislocation velocities and on dislocation densities will aid in the interpretation of high strain-rate phenomena. From the practical point of view, the effects of high-speed flow on the subsequent behavior of materials should be studied. Moreover, a phenomenological theory of the behavior of materials during high-speed forming operations should be developed.

As background for studies of the behavior of materials at high strain rates, a sound and thoroughly verified theory of plastic wave propagation is needed. Furthermore, since the energy that is dissipated by fast-moving dislocations appears to be somewhat greater than existing theoretical estimates, a major refinement in the theory may be needed to account for the observed behavior.

3.1.5. Effects of Grain Boundaries on Slip Distribution

A grain boundary (Amelinckx and Dekeyser, 1959R) can increase the resistance to deformation in two ways. First, it may serve as a barrier to dislocations generated in adjacent grains. Second, it may induce multiple slip in neighboring grains. It is important to distinguish between these two effects, since the interpretation of any particular correlation between grain size and some specific feature of the stress-strain curve will depend upon which effect is dominant. In polycrystalline materials, the relative importance of the two principal strengthening effects associated with grain boundaries is determined by several factors: the number of slip systems available, the ease of cross-slip, the presence of alloying elements, the nature of dislocation-solute atom interactions, the grain size, and the degree of plastic strain. A complex small-angle dislocation boundary network increases the flow stress primarily by providing barriers (Amelinckx and Dekeyser, 1959R). Simple dislocation boundaries cannot act as barriers unless pinned by solute atoms.

When the grain size is fairly large, multiple slip is often confined to the grain boundary regions at small strains. Under these conditions, the larger the grain size, the lower the overall strain hardening rate. This is because the fraction of the total volume

occupied by regions undergoing rapid strain hardening becomes progressively smaller the larger the grain size. As the plastic strain increases, multiple slip spreads inwards and the volume undergoing rapid strain hardening increases. Eventually, a strain hardening rate is reached which is independent of grain size. When discontinuous yielding occurs, however, other factors must be considered.

Grain boundaries appear to exert a particularly marked effect on the early portion of the stress-strain curve. The nature of this influence depends upon the plastic anisotropy of the solid, and upon the alloying or impurity elements present; however, few critical experiments have as yet been made to elucidate the effect of grain boundaries. This aspect of plastic flow could be studied effectively with polycrystalline alkali halides. At low temperatures, glide occurs in these materials on the six $\langle 110 \rangle$ $\{110\}$ systems, whereas at higher temperatures, other slip systems become operative, increasing the total number to twelve. The effect of grain size on the shape of the stress-strain curve should therefore vary markedly with temperature. Very little work has been done to investigate grain size effects as a function of slip mode for pure solids.

One of the more important areas to be investigated lies in determining the effect of grain boundaries when the grain sizes are of the same order of magnitude as those of the finer details of the work-hardened state. If the grain size could be reduced to the scale of the slip pattern (i.e., in the range from 100 to 1μ), then grain boundaries might be expected to have a profound influence on the mechanical behavior. When the grain size is as small as one micron, it may be very difficult for dislocation sources to generate complete loops, and this might increase the so-called limit of proportionality by orders of magnitude.

3.1.6. Grain Boundary Sliding and Intergranular Cracking

Existing observations on grain boundary relaxation processes, particularly those responsible for grain boundary sliding and intergranular cracking, are fraught with contradictions. Although much experimental evidence has been accumulated for reasonably pure metals and alloys, conflicting conclusions have often been drawn from the results and, at present, the phenomena involved are by no means well understood.

Since the effects in question are manifest at high temperatures, it is understandable that the principal experimental effort has been to find out to what extent they control creep behavior. The

extent to which grain boundary sliding contributes to the creep strain, and the relationship between sliding and the movement of dislocations within neighboring crystals, are problems which cannot be resolved satisfactorily until the nature of grain boundary sliding itself is clearly established. It is not known, for example, whether the shear of one grain over another is a quasi-viscous process confined within the boundary region itself or whether dislocation movements in the immediate neighborhood of the boundary are responsible for the observed translations (Grant, 1959; Gifkins, 1959R). If such zones of intense shear are produced by crystallographic slip, their width is often beyond the resolution of the optical microscope.

If the high resolution techniques currently available could be extended to permit the direct observation of events occurring at or near a boundary, this aspect of the problem could be clarified. Of the many features that have been revealed experimentally, the following appear to be the most significant: (1) the activation energy for the creep of polycrystals does not appear to be a function of grain size and corresponds to that of self-diffusion (Dorn, 1954); (2) a zone of intensive polygonization frequently develops in the immediate vicinity of the boundaries (Grant, 1959; Gifkins, 1959R); (3) under certain circumstances, the secondary creep rate is found to decrease with decreasing grain size (Parker, 1958).

3.1.7. Crack Formation at High Temperatures

Two basic types of intergranular cracking (Grant, 1959; Gifkins, 1959R) have been recognized. One takes the form of a wedge-shaped void generated by boundary shear at the common junction between three grains. Such cracks form at comparatively high stresses. The other type of intergranular cracking is commonly observed at low stresses and high temperatures and is apparently associated with the formation and growth of isolated intergranular cavities. The mechanism of this form of cracking is not fully understood, but several important features are becoming clear. For example, grain boundary sliding is generally associated with intergranular cracking. Although intergranular pores are rarely formed in the absence of grain boundary sliding, the latter does not necessarily lead to the formation of cavities. There is evidence that nucleation of cavities occurs at grain boundary irregularities such as are provided by particles of a second-phase or by the intersection of slip bands or small-angle boundaries. The fact that intergranular cracking or void formation is rarely observed

in high purity metals suggests that the introduction of trace elements, either in solid solution or as a second phase, plays an important part in restricting accommodation during boundary sliding. There is a great need for critical experiments in this area, particularly with regard to: (a) the possible correlation between the distribution of slip and small-angle boundaries and the nucleation of cavities in bi- and tricrystals, and (b) the effect of alloying on the distribution and nature of slip introduced under comparable conditions.

The growth of cavities at grain boundaries appears to demand both vacancy diffusion as well as the application of a tensile stress (Gifkins, 1959R). It has been suggested that growth occurs by a process which is, in effect, opposite to pressure-assisted sintering (i.e., hot-pressing of powders). In the absence of a tensile stress, isolated pores decrease their volume by the diffusion of vacancies to appropriate sinks (principally grain boundaries). The driving force for this process is $2\gamma/r$, where γ is the specific surface energy and r is the radius of the pore. A tensile stress greater than $2\gamma/r$ causes the cavity to grow by the same vacancy diffusion process, and the specimen to elongate in the direction of the tensile stress by the accumulation at grain boundaries of atoms removed from the surface of the cavity. This theory of crack growth seems to offer considerable promise and deserves careful evaluation. The influence of tensile stresses on closed-pore sintering using powder metallurgy techniques (particularly with transparent ionic solids) may yield valuable information regarding the kinetics of the process.

3.1.8. Twinning

Twinning, which differs considerably from slip or glide in its essential characteristics, is a fundamental mode of plastic deformation. As such, it represents an important subject for study. The twinning process is also of major interest because: 1) it has many features in common with martensitic phase transformations, e.g., the kinetics are basically similar; 2) it may have an important influence on strain-hardening; 3) twinning sometimes causes crack nucleation; and finally, 4) it may strongly affect crack propagation.

The crystallography of twinning in many systems has been worked out in detail and rationalized with considerable success. What remains quite obscure is the mechanism by which a crystal lattice becomes twinned. Dislocation theories of the mechanism have been proposed, but their validity has not been checked by

critical experiments. Some questions which need to be investigated experimentally are:

- (a) What stress is needed for homogeneous twin nucleation?
- (b) What crystal defects act as twin nucleation catalysts?
- (c) Can twinning dislocations be observed?
- (d) How fast do twins propagate parallel and perpendicular to the twinning direction?
- (e) What are the effects of dislocation networks, impurities, and temperature on twin propagation?
- (f) What is the maximum speed with which twins can propagate?
- (g) What are the energies of various mechanical twin boundaries?

3.2. Fracture Phenomena

3.2.1. Introduction

Fracture in crystalline materials is a subject of great complexity and until recently (Averbach, et al., 1959R) little progress has been made toward an understanding of the processes involved in the nucleation and growth of cracks. One reason for this has been the discouraging mass of confusing and seemingly contradictory information. Fortunately, this is one of the few areas of research in which the practical value of basic knowledge is obvious; consequently, research in this field has been well supported.

Many crystalline materials fail because they deform plastically in local regions. This local plastic flow causes stress concentrations that lead to fracture. Therefore, one of the most interesting and potentially important fields of investigation is the relationship between plastic deformation and fracture. Some important questions which have not yet been resolved are: a) how are cracks nucleated by slip? b) what determines the amount of plastic strain required for crack nucleation in any given material? c) how is the growth of a crack influenced or controlled by the plastic properties of the material?

A full understanding of the nucleation and growth of cracks in plastic crystals requires an accurate knowledge of the morphology of slip phenomena, and much still remains to be learned in this area. Most of the current theories of cleavage fracture assume that the crack is nucleated by the coalescence of edge dislocations at the head of piled-up arrays (Cottrell, 1959). Such coalescence requires high stress concentrations at barriers, e.g., grain boundaries, where dislocation pile-ups occur. The magnitude of such

stress concentrations depends on the number of dislocations in the array, the strength of the barriers, and how fast the dislocations arrive at the barrier (Cottrell, 1953R). These factors, in turn, depend on the details of the plastic flow mechanism. Grain boundaries are serious obstacles to slip only when the number of slip systems is inadequate to permit the sum of all the Burgers vectors of the dislocations piled up against both sides of the boundary to be zero.

Once a crack is nucleated, the stress level required for the growth or propagation of the crack depends primarily upon the surface energy of the material, its microstructure, and its dynamic plastic behavior (Petch, 1959; Low, 1959). The strain energy released by the growth of the crack must exceed the sum of the surface energy and plastic work involved in extending the crack. Plastic flow near the tip of a crack tends to decrease the stress concentration in the vicinity by blunting the end of the crack, and extensive local slip also results in the absorption of a large amount of energy because of plastic work (Gilman, 1957).

An understanding of certain other features of the deformation morphology is equally essential for the formulation of a satisfactory model of the fracture process. More quantitative information is needed about the stress levels required to move dislocations on various crystallographic planes, and about the temperature dependence of these stresses. Moreover, crack propagation is a dynamic process, and it is conceivable that the strain rate near the tip of an advancing crack may be orders of magnitude greater than that obtainable in most mechanical tests. Little experimental or theoretical information is now available concerning the mechanism of plastic deformation at extremely high strain rates, where the shear stresses may approach the theoretical shear strength of the material.

The various kinds of fracture, such as brittle or cleavage fracture, ductile fracture, and fatigue, should not be considered as unrelated phenomena. Plastic deformation is often necessary for the formation of crack nuclei in each case. Growing cracks frequently follow the path requiring the least surface energy; this may be along a slip band, a grain boundary, a cleavage plane, etc.

3.2.2. Cleavage Fractures

It is well known that cleavage fractures are often initiated in semibrittle materials by plastic flow (Averbach, et al., 1959R). The evidence comes from various sources, e.g., 1) the tensile frac-

ture stress for steel at low temperatures is very nearly equal to the yield stress in compression; 2) when plastic flow is suppressed in zinc crystals, the fracture stress rises to a very high value; 3) cracks in ionic crystals, such as MgO, initiate at the intersections of slip bands.

Cleavage fracture requires the presence of local tensile stresses, and several theories for converting local plastic shears into large localized tensile stresses have been proposed. These always postulate the existence of some kind of obstacle which blocks the shear, thereby creating a discontinuity in the shear strain (Cottrell, 1959R). At the discontinuity, large tensile stresses are produced. An additional requirement is that the local shearing must occur rapidly at the higher temperatures, otherwise high stresses cannot be obtained because there will be time for stress relaxation.

Another way in which cleavage fractures may sometimes be initiated is by chemical attack on the surface of a material. This process apparently depends upon the production of corrosion products which "wedge" open any chink in the surface until a crack develops that is large enough to run catastrophically.

In the absence of pre-existing cracks, or of a mechanism for crack nucleation, theory predicts that a solid should be able to withstand stresses that are at least equal to a few per cent of its elastic modulus. This has been verified for carefully prepared silicon crystals, as well as for whiskers of many materials.

Cleavage crack propagation, as contrasted with crack initiation, has received relatively little attention. It has long been known that, for an elastic body, the minimum energy required to make a crack run is the surface energy of the newly formed crack faces. For a body that is slightly plastic, deformation occurs in narrow zones along the crack surfaces and, therefore, additional energy must be provided in order to propagate the crack. It is also known that the velocity of a running crack is limited by the rate at which the stress can relax in the portions through which the crack has passed. This means that a crack cannot move faster than about one-third of the longitudinal sound velocity. Not much is known about flow and fracture at high velocities; most of the existing knowledge is qualitative in nature.

Quantitative studies of crack propagation are needed to determine the effective fracture surface energies so that these can be compared with theoretical calculations. Furthermore, the effect of such variables as crystal structure and microstructure on crack propagation also merits careful study.

3.2.3. Effect of Grain Boundaries on Cleavage

The fact that the ductile-brittle transition of b.c.c. metals containing interstitials is shifted to lower temperatures as the grain size decreases has been clearly established experimentally (Averbach, et al., 1959R). The brittle fracture stress is inversely proportional to $D^{1/2}$ where D is the average grain diameter. Theoretical treatments of the grain size effect differ largely in the assumptions which they make with regard to the influence of grain size in limiting the growth of cracks initiated by dislocation interactions. In this connection, more careful experimental study is necessary to define the role of boundaries in restricting the growth of cracks, particularly from the viewpoint of the effect of grain size (especially in the range 100μ to 1μ) on the velocity which a cleavage crack can attain as it traverses a single grain. Furthermore, since subboundaries appear to have a significant effect on the yield behavior of mild steel, whereas they should have little or no direct influence on the velocity of propagation of a crack through a grain, a study of the effect of a change in substructure size for a given grain size should provide a means of clarifying the theoretical treatment of the grain size effect.

The role of grain boundaries in the brittle fracture of ionic solids is virtually unexplored. It is of particular interest to study the effects in crystals having the rocksalt structure, because detailed knowledge of dislocation behavior and fracture mechanisms in monocrystals is rapidly being accumulated for materials of this kind.

3.2.4. Ductile-Shear Fracture

The fundamental characteristic which sets a ductile-shear crack apart from a macroscopic cleavage crack is that in order for a ductile crack to propagate, plastic work at the crack tip is mandatory. Ductile cracking occurs only as a result of localization of slip, e.g., at or within bands of intense shear. Comparatively little fundamental work has been done in connection with ductile cracking (in contrast to that associated with cleavage), but recent experiments have clarified in a qualitative way some of the important features of the process. Experimental evidence indicates that when a cavity is created in some way within the neck of a tensile specimen, it can widen or grow simply by glide on slip bands which run into or around the cavity. Furthermore, evidence suggests that the intense slip bands which form in the later stages of necking in a thick plate are the result of localization of glide by internal cavities. In cylindrical specimens, many cavities

open up within the necked region, each of which grow by local glide. Such cavities eventually coalesce to form thin lenticular-shaped cracks having a multitude of rough facets. The main fracture path is approximately parallel to the plane of maximum shear stress. This mechanism of crack growth is well established only for aluminum and copper.

One essential feature of a shear crack is that it tends to propagate parallel (macroscopically, at least) to a glide band. Materials that can maintain a high strain-hardening rate, even after appreciable strains, can fracture by this mode without necking. Several investigators have shown that in materials of this kind, the macroscopic plane of the shear crack is parallel to coarse slip bands which form just before fracture. The surface details of shear cracks in steels and in face-centered cubic alloys suggest that they propagate discontinuously, with the formation of both inclusion cavities and so-called "de-cohesion cracks," ahead of the main crack front. The de-cohesion cracks are of extreme interest because they appear to represent regions where, on a microscopic scale, a crack propagates over a crystallographic slip plane across which the cohesion has been lowered, due perhaps to a high density of dislocations. This mode of fracture has been observed in the rupture of thin films of austenitic stainless steel using high-resolution transmission electron microscopy. In certain regions, the foil separates along a zig-zag path which follows two intersecting systems of slip dislocations. It is important to determine whether there is any essential difference between cleavage on a microscopic scale and the above-mentioned de-cohesion cracks.

Too little fundamental evidence is available on which to base a satisfactory theory of shear fracture. Inclusions appear to play an important role. One fruitful experimental approach would be to study the behavior of solids containing a dispersed phase that has been added in a controlled fashion.

3.3. Flow and Fracture of Two-Phase Systems

The fundamental question of importance is an obvious one—what characteristics of the dispersed phase control the mechanical properties of the aggregate? Little systematic analysis or critical experimental work has been done to answer this question. There are many reasons for this. The behavior of single-phase polycrystals is still incompletely understood, especially when the grain size or subboundary spacing is of the order of 1–10 μ . This is precisely the scale at which dispersions exert their most inter-

esting effects. The principal experimental difficulties have been to obtain controlled microstructures which are amenable to analysis, and to be able to observe the details of slip interactions between the two phases. Nevertheless, our present knowledge and techniques have progressed to the stage where it should be possible to begin to understand the behavior of multiphase systems.

With respect to the normal behavior of a single-phase matrix, the introduction of dispersed particles of a second phase may increase the flow stress and strain hardening rate, decrease the creep rate at elevated temperature and modify the rupture or fracture characteristics at all temperatures.

Two substantially unexplored but potentially fruitful experimental approaches are open for the study of two-phase systems. One is through the use of transparent crystals containing a dispersed phase. Optical, decoration, and etch pit techniques now in use for studies of single-phase solids of this type could be gainfully employed for such investigations. The second technique is electron microscopy of thin sections. There is no doubt that this approach will prove extremely enlightening, especially for exploring the behavior of commercially useful materials.

4. ANELASTICITY AND INTERNAL FRICTION

4.1. Introduction

This discussion deals with the study of flow processes that take place at very low stress levels (Zener, 1948R, Nowick, 1953R). Two general types of relaxation phenomena are involved, one of which produces a peak in the internal friction as a function of temperature while the other does not.

Three parameters are used to describe this first type of relaxation phenomenon, namely: the relaxation strength, Δ_M (where M refers to the elastic modulus appropriate to the particular mode of deformation used), which is obtained from the area under the peak; the relaxation time, τ , obtained from the position or temperature of the peak; and the breadth of the relaxation spectrum, obtained from the width of the peak. In relaxation processes involving defect or atom movements, τ varies with temperature according to $\tau^{-1} = A \exp (-H/kT)$, where A is the appropriate frequency factor, H is the activation energy, and k the Boltzmann constant. For a process having a single time of relaxation, the width of the internal friction peak (plotted against T^{-1}) at half maximum is given by $\sigma (T^{-1}) = 2.62 (k/H)$. Static experiments, such as stress relaxation at constant strain or creep at constant stress, may also

be used to determine the activation energies but internal friction measurements generally yield more precise values.

The second type of phenomenon does not give rise to a peak in the internal friction vs temperature curve. Flow processes (e.g., creep) that do not come to a definite final state of strain with time, or static hysteresis phenomena which do not involve a flow process at all, contribute to this part of the internal friction. Usually, this type of behavior is more difficult to interpret unambiguously.

An understanding of these various phenomena must be sought in terms of the atomic mechanisms which produce them. Studies of lattice defects, atomic mobility, and phase transformations can be greatly facilitated by the use of internal friction and anelasticity experiments.

4.2. Effects Caused by Point Defects and Solute Atoms

Relaxation phenomena due to preferential ordering of atoms or lattice defects under an applied shear stress fall into this category. In order to contribute to the relaxation, a defect must be of such a character that it distorts the lattice preferentially along some particular crystallographic direction. The best known example is the Snoek (1941) relaxation due to interstitial solutes in body-centered cubic metals (Powers and Doyle, 1959). Other possibilities include relaxations due to preferential reorientation of divacancies, associated vacancy-solute pairs, and pairs of solute atoms. Snoek relaxation obeys the equations for a single relaxation time, with Δ_M directly proportional to the solute concentration, and τ^{-1} proportional to the diffusion coefficient of the interstitial solute. This effect has been used with great success to study solubility, precipitation, and diffusion of oxygen, nitrogen, and carbon in Fe, V, Nb, and Ta (Wert, 1950A, 1950B, 1950C; Wert and Zener, 1949).

Among the newer discoveries, the implications of which have not yet been fully explored, are the peaks due to hydrogen and deuterium in iron. Recent work has also demonstrated the existence of interactions between solute atoms. Thus, shifts have been found in the positions of the oxygen and nitrogen peaks in Ta due to O-O, O-N, and N-N interactions (Powers and Doyle, 1956). Some work has also been carried out on the interactions between interstitial atoms and substitutional impurities, but there is room for further investigations along these lines. A particularly intriguing thought is the possibility of producing internal friction peaks due to association of an interstitial solute with a substitu-

tional impurity in an alloy which is not of the b.c.c. type, where the interstitial solute alone would not produce a strong relaxation peak. There has been some evidence for this type of phenomenon, but a more detailed investigation would be appropriate.

In substitutional solid solutions, the best known effect is the Zener relaxation (also called the pair-reorientation effect, or the stress-induced ordering effect). It exists in a wide range of solid solutions and is describable in terms of a single time of relaxation; however, a full understanding of this effect has not yet been achieved. Zener's original idea (1947, 1948R) that pairs of solute atoms are reoriented by an applied shear stress, and LeClaire and Lomer's refinement (1954) in terms of quasi-chemical theory are unable to explain the orientation dependence recently found by Seraphim and Nowick (in press) in single crystals of Ag-Zn (f.c.c.) and Li-Mg alloys (b.c.c.). There is a need for a new theoretical approach. An improved understanding of this phenomenon should lead to improvements in alloy theory as well as to more extensive applications of Zener relaxation experiments. From the kinetic point of view, it is clear that this process involves a simple atom rearrangement, and that the activation energy, H , is very nearly equal to the activation energy for self-diffusion. Nevertheless, there are small discrepancies which must still be explained, viz., that H always appears to be lower than the self-diffusion activation energy for both the solute and the solvent.

The relaxation strength, Δ_M , is a useful indicator of structural changes which occur in solid solution alloys during such processes as ordering and precipitation, while measurements of τ are of value in studies of atom movements in the presence of an excess of vacancies, e.g., in rapidly quenched alloys (Roswell and Nowick, 1953). Additional work is called for on cold-worked and irradiated alloys. Furthermore, Zener relaxation studies should be extended to ternary systems containing a small percentage of the third component in order to obtain information on solute atom interactions and solute-vacancy interactions. Experiments should also be carried out on alloys of various compositions which exhibit long-range ordering effects.

A peak due to solute-vacancy association should be observable in some alloy systems. This peak could be useful in the studies of defects. There has been some evidence for such an effect in NaCl crystals containing impurities and it would be desirable to look for this same phenomenon in metals. A related but more complex

phenomenon due to solute clusters which include a vacancy has recently been found in duralumin.

Finally, the divacancy peak measurements would be most valuable in studying defects produced by quenching, irradiation, and deformation, because they provide a means for distinguishing single vacancies from divacancies (something that resistivity measurements, for example, cannot do). At present, considerable confusion exists in this area. Hasiguti (1953) claims to have found such a peak in various metals after cold working. However, others have looked for it unsuccessfully. Alloys such as NiAl, which have an intrinsic defect structure, are likely to show such a peak. The divacancy peak should also occur after low temperature irradiation with high energy particles.

4.3. Effects Originating at Interfaces

The grain boundary relaxation phenomenon, found in all polycrystalline metals and alloys, is of considerable interest because it is the result of boundary flow (Girkins, 1959R; Nowick, 1952R). In contrast, most mechanical studies of polycrystalline materials reveal only the effects of the boundaries in perturbing plastic flow. Unfortunately, although the existence of the grain boundary relaxation phenomenon has been known for more than a decade, it is still poorly understood. Boundary relaxation is not as distinct an effect as the processes discussed previously. The damping peak is generally two to three times as broad as that exhibited by a unique relaxation event. The phenomenon is extremely sensitive to the presence of solute atoms (even at average concentrations as low as 0.01%) because of the tendency of solute atoms to segregate in boundaries. It is also very sensitive to the presence of dissolved gases.

Recent work has shown that the peak found in pure metals is suppressed by small concentrations of impurities and, for the same frequency, it is replaced by another peak located at a higher temperature. However, the total relaxation strength apparently remains constant. These observations appear to be sufficient to rule out the theories of Mott (1948) and of Kê (1947) for this kind of relaxation, but thus far no satisfactory alternative theories have been proposed. One complication is the peculiar way in which the activation energy of the peak in the pure metals is increased by solute additions before the peak is finally suppressed. In copper, for example, the activation energy appears to increase from 40 to 80 kcal/mole as impurities are added. Recent experiments have also shown that quenching of an alloy results in a considerable

decrease in τ for the grain boundary relaxation, but the reason for this effect is not clear.

It is important to try to correlate grain boundary relaxation with grain boundary slip of bicrystals at elevated temperatures. The interrelation between the two types of measurements is not clear. The fact that the relaxation time obtained from internal friction measurements is proportional to the grain dimension (based on limited evidence) is surprising because this implies that the only pinning points for slip are the grain corners, i.e., that boundary irregularities do not play a significant role.

There is also a need for additional theoretical work in the light of the newer experiments. In particular, the relative contributions to the boundary relaxation process from slip and from migration of the boundary must be clarified.

An entirely different damping peak, apparently due to twin boundaries, has been found in a Mn-Cu alloy. Zener has suggested that it represents a true relaxation phenomenon, but the frequency dependence of the effect has never been studied. Experiments on the In-Tl system, which shows similar twin boundaries, would be of interest.

There is a high temperature "background" internal friction which increases monotonically with temperature. It is apparently due to boundaries because it is at least an order of magnitude smaller in single crystals. It has been suggested that relaxation at grain corners, where the stress builds up during the normal grain boundary relaxation, is responsible for this "background" effect; however, this explanation can only be considered tentative at best. Particularly noteworthy is the low activation energy reported for this effect, relative to the grain boundary peak. Boulanger (1949) has found a particularly large effect, apparently of this kind, with a wide range of activation energies in polygonized aluminum. Friedel, Boulanger and Crussard (1955) interpret the effect in terms of the displacement of dislocation walls in the polygonized structure by glide and diffusion. In view of the importance of such phenomena to the understanding of the high temperature behavior of materials, studies of the background internal friction should be pursued with renewed vigor

4.4. Effects Associated with Dislocations and with Interactions Between Point Defects and Dislocations

This section will be concerned with phenomena that are associated with the properties of individual dislocations rather than with dislocation arrays in grain boundaries or polygonization

boundaries. There are two distinct relaxation effects due to dislocations which appear as internal friction peaks. In addition, more complex effects have been observed which probably cannot be classified as true relaxation phenomena.

The dislocation-impurity interaction peak was discovered by Snoek (1941) in connection with his work on carbon in iron. The peak occurs at about 200°C in cold-worked iron containing carbon after the lower temperature Snoek relaxation effect has begun to disappear (Köster, Bangert and Hahn, 1954). It has been attributed to the "drag" caused by excess solute atoms concentrated at dislocations. Various complications have recently been pointed out, suggesting that previous interpretations require modification. One surprising fact is that in Fe-C there is a period in which neither peak is present, whereas for Fe-N the new peak builds up as the Snoek peak disappears. Although this phenomenon clearly appears to be related to a dislocation-impurity interaction, the details are still in doubt. One limitation has been that the effect has not been found in a wide enough range of materials. It would certainly be desirable, for example, to see if substantial solid solutions can be made to display this peak. Interestingly enough, Kessler (1957) has observed a peak in pure Ge which he attributed to vacancy-dislocation interactions.

The Bordoni peak (1954) is a relaxation phenomenon that occurs in a number of pure metals that have been cold worked. These peaks for various metals are located in the vicinity of 40°K to 100°K for frequencies in the 10 kc range. The effect is a true relaxation phenomenon, in that it has the appropriate frequency dependence. The effect is essentially strain-amplitude independent. Also, a second peak has been found on the low temperature side of the principal peak in copper (Niblett and Wilks, 1956, 1957). The peak position is surprisingly insensitive to small amounts of impurity and to the amount of deformation. Because of this, Seeger (1956) ruled out an earlier theory of Mason (1955) according to which the peak location depended on the length of free dislocation segment; instead, Seeger attributed the peak to the formation of kinks of critical size in dislocations that are confined to the principal crystallographic directions. This theory predicts a rather large Peierls stress ($\sim 10^{-3}\mu$, where μ is the shear modulus) (Bordoni, et al., 1959) as well as an amplitude dependence of the internal friction. However, this amplitude dependence has not been observed. More recent work by Thompson and Holmes (1959) has revealed that the peak has other complexities and their experiments have also shown that irradiation sufficient to eliminate most

of the room temperature damping has only a slight effect on the Bordoni peak. As a whole, Seeger's original theory has withstood these experimental tests reasonably well, although several modifications have been introduced (Donth, 1957; Seeger, Donth and Pfaff, 1957). The principal need is for further experiments on a wide range of materials. Thus far, only f.c.c. metals have been studied in any detail, with emphasis on Cu and Al. A better perspective can be obtained only after the same peak has been found and studied in such materials as b.c.c. metals (Chambers and Schultz, 1960), germanium, and the alkali halides.

The static hysteresis effect is the well-known, strongly amplitude-dependent damping first observed by Read (1940). This effect was shown to be independent of frequency, and was, therefore, attributed to a static hysteresis phenomenon. A quantitative theory, developed by Granato and Lücke (1956), has been widely used to analyze such data. The theory has been found to work well for crystals that have been doped with impurities. With such crystals, the amplitude dependence rises sharply at a strain amplitude which depends on the impurity content, and which corresponds to the stress-induced break-away of the dislocation segments from impurity atoms to which they are pinned. According to this theory, a plot of $\log(\epsilon \cdot \delta)$ versus ϵ^{-1} (where δ is the internal friction and ϵ is the strain amplitude) should give a straight line. The slope of the line should be proportional to impurity concentration while the intercept should be dependent upon the dislocation density. These predictions seem to fit the experimental data rather well. However, for relatively pure crystals, the Granato-Lücke theory does not appear to be applicable. What is needed are modifications of the theory which take into account thermal activation. This should help to explain some of the deviations observed at low amplitudes, especially in the purer crystals. In general, however, the phenomenon appears to be reasonably well understood.

There is evidence that vacancies which are quenched into a metal from a high temperature and which are allowed to diffuse to dislocations will suppress the static hysteresis damping. Effects of this kind are surely related to the quench-hardening phenomenon as observed in measurements of the yield stress. It would be valuable if further internal friction studies were made, preferably in pure Au or Cu, on rapidly quenched thin samples. The rate at which the hardening effect develops on subsequent aging at various temperatures should be studied. In this way, a better understanding of the nature of the dislocation pinning may be achieved.

The Köster effect (1940) identifies the internal friction which

appears in freshly deformed crystals in the vicinity of room temperature and which decays rapidly with time. It appears to be practically independent of frequency and strain amplitude. The Köster effect may possibly be due to the same process that is responsible for the amplitude-independent internal friction in undeformed crystals, which also appears to be independent of frequency, or very nearly so. One theory of this phenomenon is that originally suggested by Koehler (1952) and later modified by Granato and Lücke (1956). It pictures dislocation loop oscillation in the presence of a damping force, and predicts a damping which is proportional to the frequency in the usual frequency range. Some observations in the megacycle range agree with this frequency dependence, but data in the kilocycle range do not. Furthermore, essentially the same type of damping is observed at frequencies of the order of 1 cps. To resolve present inconsistencies, better measurements of the frequency dependence of the internal friction are required. It is surprising that, in spite of disagreements regarding the frequency dependence, some of the predictions of the Koehler theory are still valid, viz., the decrease in damping with decreasing dislocation loop length (e.g., through irradiation) and, more quantitatively, the apparent dependence of damping on L^4 , where L is the loop length.

The manner in which the Köster damping decreases with time after deformation is also of interest. Two theories have been suggested: the dislocation rearrangement theory (Nowick, 1955) and the theory of pinning of dislocations by the point defects created by deformation (Granato, Hikata and Lücke, 1958). The relative merits of these theories cannot be evaluated with existing evidence. According to the rearrangement theory, the effective dislocation density should decrease with time, while according to the pinning theory it is the loop length that decreases. Thompson and Holmes (1959) as well as Gordon and Nowick (1956) have shown how to determine these quantities from irradiation data, and the use of irradiation to obtain the dislocation density and loop length at any time in the recovery process should help to solve this problem (Bauer and Gordon, 1960). Experiments in which recovery is carried out under a static stress, and in which the internal friction is studied subsequent to the release of the stress have proved very interesting in the past and might well be pursued in more detail.

Other dislocation effects of interest, that have been studied in a less complete way, include the high temperature damping which has been found to vary exponentially with T^{-1} (Mason, 1958R)

and high amplitude damping and its relation to fatigue. Experiments of this kind seem to offer a promising approach to the problem of fatigue (Mason, 1958R; Thompson and Wadsworth, 1958R).

4.5. Effects Associated with Phase Transformations

Among the various types of phase transformations to be considered are order-disorder transformations, precipitation from solid solution, and transformations which involve a change in the electronic structure. The application of internal friction and related techniques to the study of such transformations constitutes essentially a virgin field.

The early work of Gorsky (1936) on the CuAu alloy revealed the existence of large anelastic effects in this system, due to the fact that ordering involves a transition from a cubic to a tetragonal structure. It would be worthwhile to reinvestigate this alloy in the light of our present understanding of anelastic phenomena.

A peak in the internal friction curve of a precipitation-hardening alloy, originating in the precipitate particles themselves, was first observed by Damask and Nowick (1955) in the Al-Ag system. The explanation given was in terms of a stress-induced change of shape of the particles due to the interaction of an applied shear stress with the internal stresses about the particles. More recently Berry (unpublished) found a peak due to the intermediate (θ') phase in Al-Cu, which is probably of similar origin. This type of phenomenon is worth investigating further, particularly over a range of compositions for Al-Ag alloys. A more quantitative theory of the effect would also be desirable.

In general, if a transformation is accompanied by dimensional changes, the application of stress will enhance or retard the transformation at temperatures near the transformation temperature. There are some examples of this in the earlier literature (Nowick, 1953R) but this type of effect has only received limited study.

5. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

5.1. Theory

1. A major barrier to progress still exists with regard to the detailed atomic structure at the center of a dislocation.
2. It is necessary to improve existing theoretical treatments of the Peierls-Nabarro force, of the separation of dislocations into partials, of the conditions for cross slip, of the force required to move a jogged dislocation, and of dislocation damping.

3. More detailed information is needed about Bordoni damping.
4. The extent to which the Frank-Read mechanism actually functions in crystals needs to be clarified.
5. The screw-screw intersection mechanism for producing vacancies has been open to doubt, and it is no longer clear how point defects are actually produced by gliding dislocations.
6. The theory and understanding of strain hardening in body-centered cubic materials have not yet been satisfactorily developed.
7. The formation, internal structure, and properties of heavy slip bands formed by fatigue need to be explained.
8. It seems doubtful whether the study of radiation-hardening will reveal any major aspects of the behavior of dislocations that could not be deduced from studies on alloys or from quenching, and work hardening experiments.
9. Further development of the theory of the coalescence of glide dislocations and their transformations into cracks is being retarded to a large extent by our inability to handle precisely problems concerning the centers of dislocations.
10. The extent to which dislocations play a direct part in ductile fractures and fractures under creep conditions is not yet clear.
11. There is need for a new theoretical approach to explain Zener relaxation phenomena (i.e., stress-induced ordering).
12. The Köster effect is in need of an improved theoretical interpretation.
13. It would be useful to extend the theory of dislocations so that it could be united with the formal mathematical theory of plasticity.

5.2. Experiment

1. The interpretation of experimental results has almost always been beclouded by a lack of knowledge of the real purity or structural perfection of specimens employed.
2. There is still no direct way to observe individual point defects or small clusters of point defects.
3. Critical experiments are needed to determine the nature of dislocation sources.
4. A satisfactory explanation of substitutional solid solution hardening is still lacking.
5. Measurement of dislocation velocities, as functions of stress and temperature, are needed in a wide variety of materials.
6. Quantitative measurements of the stresses required to make individual dislocations cut through each other, to make disloca-

tions move on cluttered glide planes, and to penetrate grain boundaries are needed.

7. The effects which specific kinds of imperfections exert on properties need to be measured quantitatively.

8. Very little is known about flow at high strain rates, and almost any systematic experiments in this field would be useful.

9. Considerable work remains to be done on the effects of grain size on properties.

10. Experiments on processes of crack nucleation should continue.

11. Crack propagation has received inadequate attention.

12. Experiments should be made to determine whether fast moving cleavage cracks can be made to propagate in face-centered cubic metals.

13. Flow and fracture in two-phase systems have received very little attention and critical and informative experiments should be carried out with such materials.

14. Zener relaxation studies should be extended to ternary systems.

15. Attempts should be made to correlate grain boundary relaxation with grain boundary slip in bicrystals at elevated temperatures.

16. The dynamics of the mechanism of twinning requires a thorough investigation on the microscopic scale.

6. REFERENCES

6.1. Reviews and General Articles

- Amelinckx, S., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 509;
- , 1959, in *Methods in Experimental Physics*, L. Marton, editor (New York, Academic Press), Vol. 6, Part A, pp. 321–356.
- , and Dekeyser, W., 1959, in *Solid State Physics*, F. Seitz and D. Turnbull, editors (New York, Academic Press), Vol. 8, pp. 327–491.
- Averbach, B. L., Felbeck, D. K., Hahn, G. T., and Thomas, D. A., editors, 1959, *Fracture* (Technology Press, MIT and John Wiley and Sons, New York).
- Brenner, S. S., 1958, in *Growth and Perfection of Crystals*, R. H. Doremus, et al., editors (New York, John Wiley and Sons), pp. 157–190.
- Cottrell, A. H., 1953, *Dislocations and Plastic Flow in Crystals* (Oxford, Clarendon Press).
- Fisher, J. C., Johnston, W. G., Thomson, R., and Vreeland, T., Jr., editors, *Dislocations and Mechanical Properties of Crystals* (New York, John Wiley and Sons).
- Friedel, J., 1956, *Les Dislocations* (Paris, Gauthier—Villars).
- Gifkins, R. C., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), p. 579.
- Hirsch, P. B., 1959, *J. Inst. Metals*, 87, 406.

- Kröner, E., 1958, *Kontinuuumstheorie der Versetzungen und Eigenspannungen* (Berlin, Springer-Verlag).
- Mason, W. P., 1958, *Physical Acoustics and the Properties of Solids* (Princeton, D. Van Nostrand Co.).
- Mott, N. F., 1960, Trans. Met. Soc. AIME, 218, 962.
- Nabarro, F. R. N., 1952, Adv. in Phys. 1, 269.
- Nowick, A. S., 1952, in *Metal Interfaces* (Cleveland, Ohio, American Society for Metals), pp. 248-267.
- , 1953, in *Progress in Metal Physics*, B. Chalmers, editor (New York, Pergamon Press), Vol. 4, pp. 1-70.
- Read, W. T., Jr., 1953, *Dislocations in Crystals* (New York, McGraw-Hill).
- Schmid, E., and Boas, W., 1938, *Kristallplastizität* (Berlin, Springer-Verlag); *Plasticity of Crystals* (translation), 1950 (London, F. A. Hughes).
- Seeger, A., 1955, in *Handbuch der Physik*, S. Flugge, editor, Vol. 7, Part 1, pp. 383-665; 1958, *ibid.*, Vol. 7, Part 2, pp. 1X210.
- Shewmon, P. G., and Zackay, V. F., editors, 1961, *Response of Metals to High Velocity Deformation* (New York, Interscience Publishers), in press.
- Shockley, W., Hoolomon, J. H., Maurer, R., and Seitz, F., editors, 1952, *Imperfections in Nearly Perfect Crystals* (New York, John Wiley and Sons).
- Thompson, N., and Wadsworth, N. J., 1958, Adv. in Phys., 7, 72.
- Zener, C., 1948, *Elasticity and Anelasticity of Metals* (Chicago, University of Chicago Press).

6.2. Original Articles

- Barrett, C. S., 1952, Trans. AIME, 188, 123.
- Bauer, C. L., and Gordon, R. B., 1960, J. Appl. Phys., 31, 945.
- Berry, B. S., unpublished.
- Blewitt, T. H., and Coltman, R. R., 1951, Phys. Rev., 82, 769.
- Boas, W., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 333.
- Bond, W. L., and Andrus, J., 1956, Phys. Rev., 101, 124.
- Bordoni, P. G., 1954, J. Acoust. Soc. Amer., 26, 495.
- , Nuovo, M., and Verdini, 1959, Nuovo Cim., 14, 273.
- Boulanger, C., 1949, Rév. Met., 46, 321.
- Chambers, R. H., and Schultz, J., 1960, Acta Met., 8, 585.
- Clarebrough, L. M., Hargreaves, M. E., and West, G. W., 1956, Phil. Mag., 1, 528.
- Cottrell, A. H., 1957a, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 509; 1957b, *ibid.*, p. 579.
- , 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons), p. 20.
- , and Stokes, R. J., 1955, Proc. Roy. Soc. (London), A233, 17.
- Damask, A. C., and Nowick, A. S., 1955, J. Appl. Phys., 26, 1165.
- Dash, W. C., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 57.
- Dash, W. C., 1958, in *Growth and Perfection of Crystals*, R. H. Doremus, et al., editors (New York, John Wiley and Sons), p. 361.
- Dehlinger, U., 1929, Ann. Phys., 2, 749.
- Donth, H., 1957, Z. Physik., 149, 111.
- Dorn, J. E., 1954, J. Mech. Phys. Solids, 3, 85.

- Edwards, E. H., Washburn, J., and Parker, E. R., 1953, *Trans. AIME*, **197**, 1525.
- Eshelby, J. D., 1949, *Proc. Phys. Soc. (London)*, **A62**, 307.
- Fisher, J. C., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 513.
- Frank, F. C., 1949, *Proc. Phys. Soc. (London)*, **A62**, 131.
- Friedel, J., Boulanger, C., and Crussard, C., 1955, *Acta Met.*, **3**, 380.
- Fullman, R. L., 1951, *J. Appl. Phys.*, **22**, 448.
- Gilman, J. J., 1957, *Trans. AIME*, **209**, 449.
- , 1959, *J. Appl. Phys.*, **30**, 1584.
- , and Johnston, W. G., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 116.
- Gordon, R. B., and Nowick, A. S., 1956, *Acta Met.*, **4**, 514.
- Gorsky, W. S., 1936, *Phys. Z. Sowjetunion*, **8**, 562.
- Granato, A., and Lücke, K., 1956, *J. Appl. Phys.*, **27**, 583.
- , Hikata, A., and Lücke, K., 1958, *Acta Met.*, **6**, 470.
- Grant, N. J., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT, and John Wiley and Sons, New York), p. 562.
- Griffith, A. A., 1921, *Phil. Trans. Roy. Soc.*, **A221**, 163.
- Hauser, F. E., Simmons, J., and Dorn, J. E., Conference on *Response of Metals to High-Velocity Deformation*, AIME, to be published.
- Hasiguti, R. R., 1953, *Proc. Int. Conf. Theor. Phys.*, Kyoto and Tokyo, p. 577.
- Heslop, J., and Petch, N. J., 1956, *Phil. Mag.*, **1**, 866.
- Hirsch, P. B., Horne, R. W., and Whelan, M. J., 1956, *Phil. Mag.*, **1**, 677.
- , ———, and ———, 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 92.
- , Partridge, P. G., and Segall, R. L., 1959, *Phil. Mag.*, **4**, 721.
- , Silcox, J., Smallman, R. E., and Westmacott, K. H., 1958, *Phil. Mag.*, **3**, 897.
- Hull, D., and Mogford, I. L., 1958, *Phil. Mag.*, **3**, 1213.
- Johnston, W. G., and Gilman, J. J., 1959, *J. Appl. Phys.*, **30**, 129; 1960, *ibid.*, **31**, 632.
- Kê, T. S., 1947, *Phys. Rev.*, **71**, 533.
- Kessler, J. O., 1957, *Phys. Rev.*, **106**, 646.
- Koehler, J. S., 1952, in *Imperfections in Nearly Perfect Crystals*, W. Shockley, et al., editors (New York, John Wiley and Sons), p. 425.
- Köster, W., 1940, *Arch. Eisenhüttenw.*, **14**, 271.
- , Bangert, L., and Hahn, R., 1954, *Arch. Eisenhüttenw.*, **25**, 569.
- Kuhlmann-Wilsdorf, D., 1958, *Phil. Mag.*, **3**, 125.
- Kunz, F. W., and Holden, A. N., 1954, *Acta Met.*, **2**, 816.
- LeClaire, A. D., and Lomer, W. M., 1954, *Asta Met.*, **2**, 731.
- Liebfried, G., and Haasen, P., 1954, *Z. Physik.*, **137**, 67.
- Low, J. R., Jr., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), p.
- Lücke, K., and Granato, A., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 425.
- , and Lange, H., 1952, *Z. Metallkunde*, **43**, 55.
- Manjoine, M. J., 1944, *J. Appl. Mechanics*, **11**, No. 4.
- Mason, W. P., 1955, *J. Acoust. Soc. Am.*, **27**, 643.

- Mott, N. F., 1948, *Proc. Phys. Soc. (London)*, **60**, 391.
- Niblett, D. H., and Wilks, J., 1956, *Phil. Mag.*, **1**, 415; 1957, *ibid.*, **2**, 1427.
- Nicholson, R. B., Thomas, G., and Nutting, J., 1959, *J. Inst. Metals*, **87**.
- Nowick, A. S., 1955, *Acta Met.*, **3**, 312.
- Orowan, E., 1934, *Z. Physik.*, **89**, 634.
- Parker, E. R., 1958, *Trans. Am. Soc. Metals*, **50**, 52.
- Peierls, R. E., 1940, *Proc. Phys. Soc. (London)*, **52**, 34.
- Petch, N. J., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), p. 54.
- Polanyi, M., 1934, *Z. Physik.*, **89**, 660.
- Powers, R. W., and Doyle, M. V., 1956, *Acta Met.*, **4**, 233.
- , and ———, 1959, *J. Appl. Phys.*, **30**, 514.
- Prandtl, L., 1928, *Z. angew. Math. u. Phys.*, **8**, 85.
- Read, T. A., 1940, *Phys. Rev.*, **58**, 871.
- Roswell, A. E., and Nowick, A. S., 1953, *J. Metals*, **5**, 1259.
- Saada, G., 1960, *Act Met.*, **8**, 200.
- Schoeck, G., and Seeger, A., 1955, *Report of a Conference on Defects in Crystalline Solids* (London, The Physical Society), p. 340.
- Seeger, A., 1954, *Z. Naturforsch.*, **9a**, 758.
- , 1956, *Phil. Mag.*, **1**, 651.
- , 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 243.
- , Donth, H., and Pfaff, F., 1957, *Discussions Faraday Soc.*, **23**, 19.
- , and Schoeck, G., 1953, *Acta Met.*, **1**, 519.
- Seraphim, D. P., and Nowick, A. S., *Acta Met.*, in the press.
- Snoek, J. L., 1941, *Physica*, **8**, 711.
- Stein, D. F., and Low, J. R., 1960, *J. Appl. Phys.*, **31**, 362.
- Taylor, G. I., 1934, *Proc. Roy. Soc. (London)*, **A145**, 362.
- Thompson, N., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), p. 354.
- Thompson, D. O., and Holmes, D. K., 1959, *J. Appl. Phys.*, **30**, 525.
- Thornton, P. R., and Hirsch, P. B., 1958, *Phil. Mag.*, **3**, 738.
- Washburn, J., 1958, in *Growth and Perfection of Crystals*, R. H. Doremus, et al., editors (New York, John Wiley and Sons), p. 342.
- Weertman, J., *Appl. Phys.*, **28**, 193; 1957b, *ibid.*, **28**, 1068.
- Wert, C., 1950a, *Trans. AIME*, **188**, 1242.
- , 1950b, in *Thermodynamics in Physical Metallurgy* (Cleveland, Ohio, American Society for Metals), p. 178.
- , 1950c, *Phys. Rev.*, **79**, 601.
- , and Zener, C., 1949, *Phys. Rev.*, **76**, 1169.
- Wood, W. A., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), p. 412.
- Zener, C., 1947, *Phys. Rev.*, **71**, 34.

Part IX

SURFACE PHENOMENA — THE NATURE AND PROPERTIES OF SOLID SURFACES AND INTERFACES

Chairman

Prof. R. GOMER
The University of Chicago
Chicago 37, Ill.

Members

Dr. F. P. BOWDEN
University of Cambridge
Cambridge, England

Prof. N. HACKERMAN
University of Texas
Austin, Tex.

Prof. PAUL HANDLER
University of Illinois
Urbana, Ill.

Prof. Dr. KARL HAUFFE
Kalle & Company AG
Wiesbaden-Biebrich
Germany

Prof. H. J. JURETSCHKE
Polytechnic Institute of
Brooklyn
Brooklyn, N. Y.

Prof. J. W. MITCHELL
University of Virginia
Charlottesville, Va.

Prof. W. D. ROBERTSON
Yale University
New Haven, Conn.

Dr. P. SHEWMON
Carnegie Institute of
Technology
Pittsburgh, Pa.

Dr. DAVID TABOR
University of Cambridge
Cambridge, England

CONTENTS

1. INTRODUCTION	Page 501
2. THERMODYNAMICS OF SURFACES	511
2.1. Basic Concepts	511

	Page
2.2. Applications of the Gibbs-Curie Criterion	512
2.2.1. Equilibrium Shapes of Small Particles	512
2.2.2. Local Equilibrium at Interface Junctions	512
2.2.3. The Gibbs-Thompson Equation	513
2.2.4. The Gibbs Adsorption Equation	513
2.2.5. Surface Tension Motivated Processes	513
2.3. Other Applications of Thermodynamics	513
2.4. Further Remarks Concerning Surface Tension	514
2.4.1. Theoretical Evaluation of E' or γ	514
2.4.2. Experimental Methods for Measuring Surface Tensions or Surface Energies	515
2.5. Summary	517
2.6. References	517
2.6.1. Reviews and General Articles	517
2.6.2. Original Articles	518
3. SURFACE STRUCTURE	518
3.1. Gross Structure	518
3.2. Atomic Structure	519
3.3. Summary	522
3.4. References	522
3.4.1. Reviews and General Articles	522
3.4.2. Original Articles	522
4. THE ROLE OF IMPERFECTIONS IN SURFACE PHENOMENA	523
4.1. Materials for Crystals	523
4.2. Growth of Crystals	524
4.3. Production of Surfaces	525
4.4. The Nature of the Surface Imperfections	526
4.5. Observations of Surface Imperfections	526
4.6. Introduction of Controlled Densities of Imperfections	528
4.7. The Role of Imperfections	529
4.8. Silver Halide Crystals as a Model System and General Conclusions	532
4.9. Structure-Dependent Chemical Activity of Metal Surfaces	533
4.9.1. Current Knowledge	534
4.9.2. Possible Areas for Future Investigation	536
4.10. References	538
4.10.1. Reviews and General Articles	538
4.10.2. Original Articles	538
5. STRUCTURE AND PROPERTIES OF GRAIN BOUNDARIES	539
5.1. Structure of Grain Boundaries	540
5.1.1. Low-Angle Grain Boundaries	540
5.1.2. High-Angle Grain Boundaries	542
5.2. Grain Boundary Adsorption of Solutes	546
5.2.1. Physical Models	546
5.2.2. Techniques for Measuring Grain Boundary Adsorption	546
5.2.3. Effects Due to Adsorption	547
5.2.4. Summary	548
5.3. Grain Boundary Mobility and Migration	549
5.4. References	550
5.4.1. Reviews and General Articles	550
5.4.2. Original Articles	550

	Page
6. FRICTION, ADHESION, AND LUBRICATION	551
6.1. Friction	551
6.1.1. Simple Adhesion Theory of Metallic Friction	551
6.1.2. Friction of Non-Metals	552
6.1.3. More Critical Theory of Metallic Friction	553
6.1.4. Friction and Chemical Decomposition	554
6.1.5. Rolling Friction	554
6.1.6. Problems in Friction	555
6.2. Adhesion	557
6.2.1. Thermodynamics of Adhesion	557
6.2.2. Adhesion of Metals	558
6.2.3. Adhesion of Contaminated Metals	558
6.2.4. Adhesion of Non-Metals	559
6.2.5. Action of Adhesives	560
6.2.6. Problems in Adhesion	560
6.3. Lubrication	560
6.3.1. "Extreme Pressure" Lubrication and "Solid-Film" Lubrication	562
6.3.2. Mixed Lubrication	562
6.3.3. Problems in Lubrication	563
6.4. References	564
6.4.1. Reviews and General Articles	564
6.4.2. Original Articles	564
7. ELECTRONIC SURFACE PROPERTIES OF METALS AND SEMICONDUCTORS	566
7.1. Electronic Surface Properties of Metals	566
7.1.1. The Experimental Situation	567
7.1.2. Theoretical Considerations	570
7.2. Electrical Properties of the Surfaces of Semiconductors and Insulators	573
7.2.1. The Cleaned Germanium Surface	574
7.2.2. The Effect of Gases on Semiconductor Surfaces	576
7.2.3. The Mobility of Charge Carriers in the Region of Space Charge	578
7.2.4. Barrier Limited Recombination at the Germanium Surface	578
7.2.5. The Surface Region	579
7.3. References	580
7.3.1. Reviews and General Articles	580
7.3.2. Original Articles	580
8. INTERACTIONS OF SURFACES WITH GASES	581
8.1. Gas-Solid Adsorption	581
8.1.1. Physiosorption	582
8.1.2. Chemisorption	584
8.1.3. Chemisorption on Metals	585
8.1.4. Chemisorption on Semiconductors	588
8.2. Heterogeneous Catalysis	591
8.3. Gas-Solid Reactions	595
8.4. Impact Phenomena	599
8.4.1. Thermal Accommodation	599
8.4.2. Condensation	601
8.4.3. Dissociation and Recombination	602
8.4.4. Sputtering	604
8.5. Modern Experimental Techniques	605
8.5.1. Ultra-High Vacuum Techniques	605

	Page
8.5.2. Cleaning of Surfaces.....	606
8.5.3. Field Emission Microscopy.....	607
8.5.4. Field Ion Microscopy.....	608
8.5.5. Field Desorption.....	608
8.5.6. Flash Filament Techniques.....	609
8.5.7. Contact Potential Determinations.....	609
8.5.8. Low-Energy Electron Diffraction.....	610
8.5.9. Infra-Red Spectroscopy.....	610
8.5.10. Electron and Nuclear Resonance Methods.....	610
8.6. References.....	611
8.6.1. Reviews and General Articles.....	611
8.6.2. Original Articles.....	612
9. LIQUID ELECTRODE REACTIONS AND CORROSION.....	613
9.1. Electrode Processes.....	613
9.2. Metallic Corrosion.....	622
9.3. References.....	624
9.3.1. Reviews and General Articles.....	624
9.3.2. Original Articles.....	625
10. A SUMMARY OF UNSOLVED PROBLEMS.....	626
10.1. Structure.....	626
10.2. Energetics.....	627
10.3. Reactivity.....	627
10.4. Conclusions.....	627

SURFACE PHENOMENA — THE NATURE AND PROPERTIES OF SOLID SURFACES AND INTERFACES

1. INTRODUCTION

Enrico Fermi once said that surfaces were very interesting but that there was so little of them. If anything, this remark emphasizes the astonishing role played by surface phenomena in science and technology. Electron emission from metal or oxide surfaces makes possible the vacuum tube, the electron microscope and the television commercial. Atomic submarine and commercial reactors depend on the successful inhibition of corrosion. Without overcoming or taking advantage of friction our civilization would be helplessly immobilized. Ballistic missile and space-ship are intimately concerned with the interaction of surfaces with molecules and ions. Gasoline, lubricating oil—the vital fluids of our age—are produced by catalytic surface reactions.

The technological importance of surface phenomena is matched by intense scientific interest in the nature and properties of surfaces and interfaces. The bulk behavior of solids, for example, is often greatly influenced by the condition of the surface or by the presence of internal boundaries, and hence an understanding of surface phenomena is frequently essential to an understanding of the bulk properties. On the other hand, most surface phenomena

owe their existence to special structures and electronic distributions which are possible only at surfaces, and these phenomena may legitimately claim our interest for this reason alone.

The realm of surface physics and chemistry is extremely broad, encompassing an extraordinary variety and diversity of surface phenomena. From this broad scientific province, we have selected for discussion the following major topics: the thermodynamics of surfaces, the structure of surfaces, the role of structural imperfections in surface phenomena, the structure and properties of grain boundaries in solids, the interactions of gases with surfaces, friction, adhesion and lubrication, electrode processes and corrosion phenomena. This list by no means exhausts the available subject material, nor does it touch comprehensively on all of the fascinating surface phenomena which have been discovered and studied in the past. It does include, however, most of the important subjects which are common to all aspects of the physics and chemistry of surfaces.

Separate sections are devoted to each of the topics listed above. While there are certain obvious connections and interrelations between these various phenomena, others may not be quite so evident. For this reason, we shall make some attempt, at the outset, to emphasize the nature of these interrelations.

It may be appropriate at this stage to define what is meant by a surface. For the purposes of this discussion, we shall regard a surface as a region of discontinuity between bulk phases (or merely bulk orientations, as in the case of grain boundaries in pure metals). This definition has purposely been made elastic or flexible, since the utility of a more rigorous specification depends to a large extent on the nature of the phenomenon under consideration. Thus, the electronic properties of semiconductor surfaces are governed by the behavior within a surface region 1000\AA or more in depth, whereas only one or two atom layers at the surface are involved in chemisorption or in characterizing the electronic surface properties of metals. Moreover, in thermodynamic treatments of surfaces, it is often convenient, but not essential, to define a surface as a geometrical two-dimensional entity.

It should be clear, therefore, that a rigorous general definition of a surface is, if not impossible, at least undesirable. Basically, this arises from the fact that surfaces never occur as isolated objects but are always intimately connected with and influenced by the phases on either side. The extent of this influence determines, in part, what must be included in a description of the "surface." In summary, then, we shall think of a surface as a

three-dimensional region, one dimension being generally much smaller than the other two.

It may also be helpful to classify the contents of this chapter according to some broad or general scheme. There are, of course, many ways in which this could be done. Based primarily on the vantage point from which each topic may be most profitably studied and discussed, we have chosen to classify the various properties of surfaces under three main categories, namely, macroscopic, atomistic and electronic.

In principle, it should be possible to predict the complete range of behavior and properties of surfaces from quantum mechanics and Coulomb's law, i.e., from a highly microscopic point of view. This is impossible in practice, and perhaps not altogether as desirable as it might appear. There are, for example, many phenomena which we might still prefer to treat at the atomistic or macroscopic level even if they could be handled quantum mechanically. Accordingly, the classification given below seems a fairly logical compromise:

MACROSCOPIC

The Thermodynamics of Surfaces

ATOMISTIC

The Structure of Surfaces

Structural Imperfections and Their Role in Surface Phenomena

The Structure and Properties of Grain Boundaries in Solids

Friction, Adhesion and Lubrication

ELECTRONIC

The Electronic Surface Properties of Metals and Semiconductors

The Interactions Between Gases and Surfaces

Electrode Processes and Corrosion

This particular classification scheme undoubtedly severs as many connections as it preserves. The same would be true, however, for any other grouping, because most of the individual topics must actually be considered from more than one point of view, and because the various subjects are intimately related and cross-linked, as the subsequent discussion will attempt to show.

It is fitting to begin with a brief consideration of the thermodynamics of surfaces since the viewpoint presented by thermodynamics is, strictly speaking, neither macroscopic nor microscopic

but quite general. Thermodynamic arguments may be usefully employed in dealing with all types of surface phenomena, for they impose the bounds within which real events must occur. Thus, the course or progress of any surface reaction is dictated by the requirement that it lead to a decrease in the total free energy of the system, while the equilibrium shapes of crystals are those for which the total free energy is a minimum. Thermodynamics has, in fact, universal appeal, and finds its applications in all branches of surface chemistry and physics.

There are many phenomena which we understand qualitatively in atomistic or electronic terms, but whose quantitative treatment at these levels is still prohibitively difficult. In these situations, thermodynamic arguments are often able to provide valuable insight into the nature of the phenomenon involved. It may be worthwhile to discuss one particularly illuminating example in some detail. An expression for the current density of thermionic emission from metal surfaces can be formulated in terms of the electron flux arriving at the emitting surface from the interior, a reflection coefficient, and the work function of the surface. To evaluate each of these terms theoretically would require, at the very least, a detailed understanding of the electron masses, energy levels and electronic motion in the metal. Such an understanding is lacking at present and hence a rigorous calculation cannot be made. Instead, a highly simplified model is generally employed and from such models the well-known Richardson equation is obtained, in apparently remarkable agreement with experiment.

Now it is possible to derive this same equation on thermodynamic grounds by utilizing the fact that, under equilibrium conditions, the rates of electron emission and absorption at the surface must be equal, as must the chemical potentials of the electrons inside and outside the metal. The thermodynamic derivation involves only the simplest gas kinetic laws together with a reflection coefficient. This treatment shows, therefore, that the final result cannot be dependent upon the details of any assumed electronic motion within the metal (except insofar as these affect the reflection coefficient, which is quite small). In turn, this serves the valuable function of preventing useless labor in more refined calculations and also serves to emphasize the limited significance of the agreement between theory and experiment.

We next turn our attention to the structure of surfaces—to the gross macroscopic features as well as the detailed atomic configuration at the surface. All surface properties are profoundly influenced by the surface structure at one or more levels. Thus,

friction and adhesion depend not only upon the macroscopic roughness of the surface (on a scale of the order of 10^{-4} cm or coarser) but also upon the atomic and electronic configurations in the regions of contact between two surfaces. In general, knowledge and control of the macroscopic surface structure and crystallographic orientation are necessary before valid conclusions can be drawn from any experiment aimed at probing deeply into the relationships between atomic or electronic structure and surface properties.

At the more microscopic level, there is no doubt that the crystallographic and atomic structure exert a profound effect on most electronic surface properties and on the properties of grain and interphase boundaries in solids. The positions of the atomic nuclei determine, for example, the positive charge distribution at the surface and this has a corresponding effect on the electronic charge distribution. Thus, the work function and the surface energy (both of which are essentially electronic properties) are influenced by the atomic configuration at the surface. Adsorption, which frequently involves an electron transfer process, is influenced by structure to the extent that it is possible to predict qualitative behavior on the basis of pairwise interactions between ad-atoms and nearest or next nearest neighbors. The initial stages of surface reactions can be treated in similar fashion.

To a first approximation, the details of atomic matching at grain boundaries in single component systems, or at interphase boundaries in multi-component systems, control the gross structural features and hence the properties of the boundary region. Thus, the presence of atomic matching or the lack thereof may determine, at least in part, whether or not a continuous, coherent oxide film will be formed on the surface of a metal during oxidation. The detailed atomic configurations at grain boundaries in metals also effectively determine the mechanism and kinetics of all processes which occur at grain boundaries, such as diffusion, impurity segregation, nucleation of precipitates, stress-corrosion cracking, etc.

Except possibly for the most closely-packed (i.e., lowest energy) crystal faces, and then only in exceedingly rare instances, the surfaces of real crystals are never smooth on an atomic scale, even when gross structural imperfections are absent. Instead, surface imperfections of various types are almost always present. These may be absorbed impurity atoms, missing lattice sites or vacancies, partially completed layers of atoms, i.e., terraces, kink steps along the edges of such terraces, steps associated with dislocations emergent at the surface, disturbances in the vicinity of grain

boundaries which intersect the surface, etc. The presence of any of these imperfections effectively alters the local atomic configuration and therefore has a marked effect on the structure-dependent behavior of surfaces.

It should be evident that the ability to determine or characterize the surface structure in its finest possible detail becomes a matter of vital importance in all surface investigations. For this reason, we have given considerable emphasis to techniques for studying the structure of surfaces. The experimental problems are, of course, formidable, but here again, the interplay between various portions of the field is impressive. In addition to primary or direct methods such as interferometric or diffraction techniques, processes occurring on surfaces frequently yield clues to the structural details. Thus, the atomic arrangement on field emission tips can not only be studied directly with the field ion emission microscope but the surface structure can also be inferred from surface mobility experiments.

The next topic which we shall discuss at some length is the role of structural imperfections in surface phenomena and the structure-dependent chemical reactivity of surfaces. This is an area which has so far received relatively little attention but is obviously of great importance. However, with the techniques now being developed for the growth of dislocation-free crystals, for the preparation of clean, well-characterized surfaces, and for the observation of structural imperfections in crystal surfaces, rapid progress may be possible in the near future. Imperfections at the surface frequently provide preferred sites for chemical reaction, e.g., nuclei for condensation or precipitation, or favorable sites for adsorption. In some processes, such as adsorption, and in certain catalytic reactions, the number density of imperfection determines their role. On the other hand, there are processes which may be initiated by a single imperfection, e.g., the growth of crystals from solution at low supersaturations by a screw dislocation mechanism, or the production of etch pits at points where dislocations emerge at the surface. As yet, however, few materials have been obtained sufficiently free from dislocations and other structural defects to permit a highly detailed study of the effects of individual imperfections on surface reactivity.

We next consider our present state of understanding of the nature and properties of grain boundaries in solids. The problems presented here are again largely structural in origin. If we knew the detailed atomic (and electronic) configurations of various types of boundaries, we could undoubtedly explain many of their

properties. It has been well established that the structure of low-angle grain boundaries in metals (and covalently-bonded materials such as germanium) can be described in terms of arrays of dislocations, and the properties of such boundaries can be accounted for satisfactorily on this basis. However, an adequate model for the structure of high-angle boundaries has not yet been developed. Since a mismatching of crystal lattices is involved, we can say, in a general sense, that the boundary region must contain imperfections, but the nature of these imperfections and the manner in which they are distributed along the boundary are not known in any detail. Even less is known about grain boundaries in ionic solids and comparatively little information of a detailed nature is available concerning the structure of inter-phase boundaries in multi-component systems of all kinds.

The general concept of the grain boundary as a region of local disorder, having a relatively high density of imperfections, provides a qualitative explanation for the enhanced mobility and chemical reactivity so frequently observed at grain boundaries, as well as for the preferential segregation or adsorption of impurities. It is difficult to measure and perhaps even more difficult to calculate absolute free surface or grain boundary energies, and reliable data of this kind are almost entirely lacking. Surface energies are important, however, in determining the equilibrium shapes of grains and in thermodynamic treatments of the adsorption of impurities at grain boundaries. A full understanding of grain boundary phenomena may therefore require a detailed consideration of the thermodynamic and kinetic factors as well as the purely structural ones.

While the subjects of friction, adhesion, and lubrication have been filed in the atomistic drawer, it should be quite clear that they involve almost every field of surface physics and other disciplines as well. Friction is governed not only by structure, but also by the presence and properties of superficial oxide films or adsorbed layers, and the same is true of adhesion. While some modes of lubrication can be explained in hydrodynamic terms, others seem to depend for their stability on the formation of adsorbed layers or on surface reactions. In addition to the semi-accidental connection between adhesion and those adsorption phenomena which make surfaces what they are, there is a more fundamental one as well. Chemisorption and adhesion both represent attempts to lower the surface free energy and therefore constitute a type of chemical binding.

One of the most fascinating and fundamental problems of sur-

face physics involves the intrinsic electronic properties of surfaces. By this we mean the electric potential, electronic wave functions, energy levels, and statistics of electrons at surfaces. A knowledge or at least a qualitative understanding of these basic quantities is prerequisite to an understanding of more complex phenomena such as adsorption and surface reactions. For example, the shape and magnitude of the electrical potential at the surface determine, in part, whether adsorbed atoms will be bound covalently or ionically to the substrate. In other words, even the grossest features of adsorption are intimately dependent on the properties of the "pure" surface. Extrapolations based on the properties of the ideally clean or pure surface are particularly useful in treating adsorption phenomena since these often represent only a relatively small perturbation.

The intrinsic electronic properties also serve to fix, in principle, a number of other quantities. For example, the surface energy of metals is basically an electronic property, since the only terms in the Hamiltonian of a finite crystal not present in that of the infinite one are electronic.

The connection between the electronic properties of surfaces and their atomic or crystallographic structure has already been mentioned. The configuration of ion cores at the surface determines the distribution of positive charge and hence influences the negative charge distribution. This manifests itself, for instance, in a variation of the work function with surface structure or orientation. Since we are unable, at present, to solve rigorously even very much simplified wave equations for structureless surfaces, phenomena such as these can be treated only qualitatively at the moment.

In the preceding discussion it has been pointed out that adsorption phenomena are basically of electronic origin. This is true even for physical adsorption, which involves a relatively mild distortion of the electron density distributions of the substrate and adsorbate, and even more so for chemisorption, which involves actual electron sharing or transfer. If more were known about the details of the available energy levels for electrons and holes at surfaces, fairly reasonable semi-quantitative calculations of adsorption energies and contact potentials could probably be made. As it is, existing correlations between the electronic properties of the substrate and its adsorption characteristics are at best very qualitative. In fact, it is often more illuminating to treat relative effects, such as the change in adsorption behavior with surface structure, in terms of nearest and next-nearest neighbor

interactions, as if pairwise bonds with the surface atoms were involved. This merely illustrates that our knowledge of the structure dependence of the electronic properties of surfaces is very rudimentary.

Adsorption invariably modifies the electronic properties of the substrate to a greater or lesser extent. Thus, adsorption on semiconductors results in the formation of exhaustion or inundation layers, which may drastically alter the electronic surface properties; even on metals, adsorption normally produces, at the very least, a change in the work function. This phenomenon may be used to advantage in studying the nature and properties of surfaces.

There is a rather close connection between adsorption and thermodynamics. To begin with, adsorption is governed by thermodynamic requirements, and the resultant changes in surface tension can be found from the Gibbs adsorption equation. Whenever adsorption is reversible, and in particular when the substrate is effectively inert, equilibrium coverages as well as rates of adsorption or desorption can be treated by thermodynamic methods, employing energy and entropy changes that can be assigned wholly to the adsorbate.

Adsorption is of interest in connection with the study of virtually all surface phenomena because it occurs so readily that, unless adequate precautions are taken to prevent it, the experiment itself may be vitiated. Without doubt, adsorption introduces some of the greatest practical problems encountered in all phases of surface study.

Adsorption is not only the usual precursor of surface reactions, but often an intrinsic part of them. Catalytic reactions occur via adsorbed intermediates so that the study of such reactions often reduces to an investigation of fairly complex adsorption processes. The surface may serve merely to dissociate the reactants or it may participate more directly by acting as a temporary electron donor or acceptor. If electron transfer occurs, even on a temporary basis, the electronic properties of the substrate become of direct significance.

Electrochemical reactions also proceed via adsorbed intermediates in many cases; in others, the presence of adsorbed layers may actually inhibit reaction. Thus, there is a close connection between electrode reactions and adsorption, although well-coordinated studies of these two phenomena are woefully lacking.

Qualitative evidence is available which indicates that the initial sites for chemical attack in corrosion processes may be associated

with imperfections of one type or another, e.g., etch pits formed at dislocations. A rather strong connection may therefore exist between surface structure and the mechanism and kinetics of corrosion reactions. However, relatively little is known at present about even such simple matters as the variation in reaction rate as a function of crystal orientation and the reasons for these variations. Clearly, a great deal of work is needed in this area.

Gas-solid reactions, such as oxidation, are always initiated by adsorption, although the character of the initial surface and that produced by reaction generally differ. It seems fairly clear, at present, that the mechanisms of oxidation of pure metals vary widely depending upon a number of conditions. When the oxide film is very thin, (i.e., in the field transport region), the properties of the underlying metal surface are of some importance insofar as they control the rate of transfer of metal ions from the substrate to the film. However, for dense, thick, adherent films, diffusion in the (semiconducting) oxide layer generally becomes rate-controlling, with the defect structure, the electronic properties and the mechanical properties of the product oxide now assuming greatest importance. The oxidation of alloys introduces still further complexities which can be understood only by careful consideration of many factors involved.

The morphology of the reaction products formed on the surfaces of metals during the initial stages of reaction are of interest for several reasons. The mechanism and kinetics of the subsequent reaction will be dependent, at least in part, on whether or not a smooth, continuous, adherent film is formed during the initial period of reaction. Under certain conditions, oxidation leads to the formation of discrete oxide nuclei rather than continuous films, or to the formation of whiskers or sheet-like protrusions growing outward from the interface. The morphology of the reaction product may be influenced by the structure of the metal surface and particularly by the presence of imperfections which may act as preferred sites for nucleation. Most of these phenomena are not very well understood at present.

In summary then, it is clear that the province of surface phenomena embraces a large number of separate topics, but that there are many connections, interrelations, and common basic problems. It would be less than honest to close this discussion without at least mentioning one other problem common to this field—the formidable experimental difficulties connected with doing meaningful work in any of its branches. These arise not only in the preparation and initial maintenance of clean, nearly

perfect, or at least adequately characterized surfaces, but also from the fact that the progress of many experiments may change the surface structure and properties in little understood ways. The classic definition, for example, excludes permanent changes in a catalyst. In practice, however, this definition often does not apply until the surface has been fashioned into a "working" catalyst by some suitable (but frequently unknown) reaction, the surface of the working catalyst being quite different from that initially prepared and characterized.

It is rather remarkable that so much of a positive nature is known about surfaces in spite of all these difficulties.

2. THERMODYNAMICS OF SURFACES

It is the great strength and fundamental weakness of thermodynamics not to inquire into the "nature of things". As a result, conclusions based on thermodynamic reasoning may become valid before a given phenomenon is understood in atomistic terms, and remain so even after detailed theories have been discarded. The foundations of thermodynamics were laid by Gibbs (*Collected Works*, 1928) more than fifty years ago. In more recent years, attempts have been made to combine, more intimately, thermodynamic and atomistic (or electronic) considerations, and to extend the treatment to non-equilibrium situations. Since several reviews on surface thermodynamics are available (MacDougall, 1939R; Gomer 1957R), including a particularly excellent summary article by Herring (1953R), this section will be relatively brief.

2.1. Basic Concepts

Although we shall regard surfaces as three-dimensional regions or phases, it is often convenient, for thermodynamic purposes, to define surface amounts in the manner of Gibbs: a dividing surface between two bulk phases is imagined. Surface quantities are then defined as the excess concentration, energy, etc. in the system relative to that which the system would possess if the bulk phases continued unchanged to the dividing surface. In recent years, an explicit three-dimensional allocation of space to the surface region has also been used by some workers (Guggenheim, 1957R). Unequivocal assignments of surface energy, entropy, etc., and of the corresponding densities (as amounts per unit surface area or surface volume) can be made on the basis of either scheme.

One of the most important quantities involved in the thermodynamic treatment of surfaces is the surface tension, γ , which is defined as the work done on the system in the creation of unit

surface area A . It can be shown that the surface tension is equal to the surface density of free energy when the surface concentrations of all constituents of the system are zero. In general, this condition can be satisfied only for single component systems (e.g., clean surfaces of pure metals). The importance of γ arises from the fact that, for stable equilibrium, $\int \gamma dA$, the integral of γ over all surfaces of the system must be a minimum with respect to all changes in form of the interface at constant volume of the bulk phases, constant chemical potentials and constant temperature, and with respect to all changes in composition of adsorbed layers which satisfy the same conditions. From this equilibrium requirement (the Gibbs-Curie criterion) it is possible to derive the equilibrium shapes of small particles, the relations between the angles of intersecting interfaces (e.g., the dihedral angles between intersecting grain or interphase boundaries in solids), the dependence of vapor pressure on radius of curvature, and the Gibbs adsorption equation, which relates changes in γ to changes in composition in the surface layer.

2.2. Applications of the Gibbs-Curie Criterion

Some of the consequences and applications of this equilibrium condition will now be examined very briefly.

2.2.1. Equilibrium Shapes of Small Particles

The equilibrium surface configuration of small particles can be derived from the so-called Wulff polar diagram in which γ is plotted as a function of the direction of the unit normal to the crystal surface. Cusps appear in the diagram corresponding to the most stable surface orientations (Herring, 1953R). This is a matter of some importance in connection with the theory of crystal growth (see Part VII) and in predicting the stability of thermal etch patterns on surfaces.

2.2.2. Local Equilibrium at Interface Junctions

If two or three grain or interphase boundaries meet along a common edge, the boundaries between the grains will reach local equilibrium at angles which are determined by the respective surface (here often called interfacial) tensions and their derivatives with respect to crystallographic orientation (Fisher and Dunn, 1952R; Smith, 1948, 1952R; Hess, 1952R). Thus, grain boundary angles can be predicted or, conversely, relative interfacial tensions can be deduced if the dihedral angles between intersecting boundaries are measured. In this manner, criteria

for wetting can also be developed. The importance of wetting in adhesion, e.g., of glass or ceramics to metals, in flotation, and in a host of other industrially important processes is, of course, quite well known.

2.2.3. The Gibbs-Thompson Equation

The vapor pressure of a small crystal (or liquid droplet) is dependent upon its size or radius of curvature and its surface tension, as expressed by the so-called Gibbs-Thompson equation (MacDougall, 1939R; Herring, 1953R). Apart from the more obvious applications of this relation, it is also of significance in connection with nucleation problems of various kinds (Hess, 1952R; Hardy and Heal, 1956R).

2.2.4. The Gibbs Adsorption Equation

The precise relationship which exists between the surface tension and the surface concentrations is embodied in the Gibbs adsorption equation (Herring, 1953R). This relation is applicable, at least in principle, to the preferential segregation of impurities at grain boundaries in solids, and it may be used to predict the changes in γ associated with the adsorption of gases at free surfaces. Since adsorption generally enhances any anisotropies in γ (as a function of crystallographic orientation), the Gibbs adsorption equation is useful in evaluating the effect of adsorption on the equilibrium shapes of crystals and on the equilibrium surface configurations developed by thermal etching.

2.2.5. Surface Tension Motivated Processes

One of the best known examples of surface tension motivated processes is sintering, a process through which the system attempts to lower $\int \gamma \, dA$ by decreasing its surface area. It is possible, in principle at least, to calculate the rates of such processes from the pertinent diffusion coefficients (surface or bulk) and the gradients of the chemical potentials of the diffusing substances (Herring, 1953R).

2.3. Other Applications of Thermodynamics

Most thermodynamic methods make use of the fact that the chemical potential, μ , of a given species must be the same throughout all phases of the system at equilibrium. Explicit use of this equality can be made in a number of ways. Thus, by combining Boltzmann or Fermi statistics for electrons or holes with the

constancy of the chemical potentials for both species, it is possible to treat the surface regions of semiconductors by essentially thermodynamic methods. Another example, i.e., the derivation of the Richardson equation governing thermionic emission from metal surfaces, has been mentioned previously.

2.4. Further Remarks Concerning Surface Tension

From even such limited considerations, it should be apparent that the surface tension is an extremely important quantity in most problems involving the thermodynamics of surfaces. The characteristics of the surface tension and methods for measuring surface tension or surface energies, particularly of solids, therefore merit further discussion. The surface tension is isotropic for liquids, but for crystalline solids it is dependent upon structure and hence is orientation dependent. Virtually nothing is known at present, either experimentally or theoretically, about the magnitude of the anisotropies involved. Herring (1953R) has shown, on the basis of atomistic considerations, that the Wulff polar diagram exhibits a point cusp whenever all three Miller indices are rational, and a knife-edge cusp for all orientations having a rational ratio of any two Miller indices, provided the temperature is so low that thermal disordering of the ideal surface structure can be neglected. As Cabrera (1959) has pointed out, this condition can be fulfilled at ordinary temperatures only for close-packed crystal faces, the condition being that $kT < E^*$ per surface atom, where E^* is the surface energy, k the Boltzmann constant and T the absolute temperature.

2.4.1. Theoretical Evaluation of E^* or γ

The surface tension, γ , is related to the surface energy, E^* , by the expression

$$E^* = \gamma - T \frac{\partial \gamma}{\partial T}, \quad (2.1)$$

and hence these two quantities are equal at 0°K. In most theoretical treatments, an attempt is made to calculate E^* rather than γ , E^* being relatively insensitive to temperature. Some estimates of the surface entropies indicate that the difference between E^* and γ at 300°K is of the order of 10%, but reliable theoretical evaluations of $\partial\gamma/\partial T$ are lacking.

For nonpolar, nonmetallic liquids, E^* can be evaluated reasonably accurately by expressing the energy of nearest-neighbor interactions in terms of the heat of vaporization, and taking into

account the number of nearest-neighbor bonds which are missing for atoms located in the surface as compared to the atoms in the bulk (Skapski, 1948; Shuttleworth, 1949). This semi-empirical method fails for metals where electronic effects tend to reduce the surface energy well below the calculated values (Oriani, 1950; Herring, 1952R). For ionic solids, lattice sum methods can be used to estimate the electrostatic (i.e., major) contributions to E^s (Ewald and Juretschke, 1953R), but again such treatments break down when applied to metals.

Although it can be shown that E^s arises solely from electronic terms in the total Hamiltonian of the finite crystal, this result does not carry us much further. The application of the modern electron theory of solids to the calculation of surface properties is still in a highly rudimentary stage (Herring, 1952R, Ewald and Juretschke, 1953R). Thus, the difficulties involved in computing an adequate potential at the surface for even an imaginary, wholly structureless metal are so formidable that little progress has yet been made using a wave-mechanical approach.

It should not prove surprising, therefore, that almost no theoretical work has yet been carried out on the structure dependence of E^s or γ . Schemes based on pairwise interactions have been used (Stranski, 1939); these are probably adequate for ionic solids, but for metals they are of qualitative usefulness at best.

2.4.2. Experimental Methods for Measuring Surface Tensions or Surface Energies

It would be pleasant if the experimental situation offered a vigorous contrast to the theoretical one. Unfortunately, this is far from true. It is a fairly simple matter to measure the surface tension of liquids at ordinary temperatures using standard methods (Adams, 1941), although even here the effects of adsorption are usually neglected or ignored. Then too, the experimental problems become much more severe when measurements are attempted at high temperatures, e.g., with liquid metals (Kozakevitch and Urbain, 1958). Even greater difficulties are encountered in making accurate measurements of the surface tensions or absolute surface energies of solids, and all presently known methods suffer from various drawbacks or limitations, as the following discussion will serve to emphasize.

1. *Cleavage Methods.* This technique is restricted to essentially brittle materials or to situations in which the energy of deformation associated with the initiation or propagation of the cleavage crack is negligible. It has been applied to materials such as mica

(Bailey, 1957), where cleavage can be carried out reversibly, and to other solids having well-defined cleavage planes (Gilman, 1959). For such special orientations, the method does yield fairly reliable values for E^s .

2. *Sintering Experiments.* As indicated previously, the driving force for the sintering process is the reduction in surface energy of the system. For simple geometries, therefore, estimates of γ can be obtained from measurements of the sintering rates (Herring, 1953R). However, the method is crude and provides only average values of γ at best.

3. *Wire Pulling Experiments.* At temperatures near the melting point, the elongation by creep of a weighted wire or thin foil is opposed by a surface tension motivated resistance to an increase in the surface area. At equilibrium, the forces due to the applied load and to surface tension are equal, and γ can therefore be evaluated under suitable conditions (Udin, 1952R; Shaler, 1953R; Pranatis and Pound, 1955). The technique is difficult in practice and there is considerable doubt whether the surfaces were initially clean or remained so in the experiments reported to date. Moreover, almost all measurements carried out so far have been made on polycrystalline wires or foils and hence the measurements yield only average values which are dependent to some extent on the particular orientations exposed at the surface.

Despite these shortcomings, it is our opinion that this method is potentially one of the best available. There does not seem to be any inherent reason why the method could not be refined by employing modern ultra-high vacuum techniques, using high purity single crystal foils of selected orientation having initially clean surfaces.

4. *Heats of Solution.* The heat of solution contains first- and higher-order terms involving the surface area. Consequently, it is possible to obtain average values of E^s and γ by comparing the heats of solution of finely divided and bulk samples, if the total surface area of these samples can be measured (Jura and Garland, 1952).

5. *Surface Energy of Liquids.* The surface energy of most liquids differs only slightly from that of the corresponding solid near the melting point. Thus, measurements of E^s or γ for liquids can be used to estimate values for solids.

6. *Measurement of Interfacial Angles.* Relative liquid-liquid, liquid-solid, or solid-solid interfacial tensions can be deduced by measuring the dihedral angles (Riegger and Van Vlack, in press) formed between three interfaces meeting along a common

line (Fisher and Dunn, 1952R; Udin, 1952R; Hess, 1952R), provided local equilibrium has been achieved at the interface junction. A variant of this general technique, known as the capillary-cone method, has recently been developed by Skapski et al. (1957). Since equilibrium is difficult to obtain in solids except at temperatures near the melting point, such methods are of limited usefulness. Moreover, measurements of dihedral angles generally yield only relative rather than absolute interfacial tensions. Nevertheless, the method has proved useful in studying the orientation dependence of grain boundary energies in single component systems, and it has also provided data on the energies of interphase boundaries in alloy systems.

2.5. Summary

Thermodynamics provides a useful tool in treating or analyzing many rate processes which are motivated by surface tension and in predicting the equilibrium surface configurations of small particles, the shapes of grains in solids, etc. In almost all applications of thermodynamics to surface phenomena, knowledge of the surface tension or surface energy is required. Experimental data on the absolute surface energies of solids is totally lacking for most materials and even for the relatively few materials which have been studied, the surface energies or interfacial tensions are not known very reliably or with sufficient accuracy. The same is true for both the temperature and orientation dependence of the surface tensions. The theoretical situation hardly appears more promising, especially for metals.

2.6. References

2.6.1. Reviews and General Articles

- Adams, N. K., 1941, in *The Physics and Chemistry of Surfaces* (Oxford, Oxford University Press), p. 362.
- Ewald, P. P., and Juretschke, H., 1953, in *Structure and Properties of Solid Surfaces*, Gomer, R., and Smith, C. S., editors (Chicago, University of Chicago Press), pp. 82-119.
- Fisher, J. C., and Dunn, C. G., 1952, in *Imperfections in Nearly Perfect Crystals*, Shockley, W., et al., editors (New York, John Wiley and Sons), pp. 317-351.
- Gibbs, J. W., 1928, *Collected Works* (New York, Longmans, Green and Company).
- Gomer, R., 1957, in *Science of Engineering Materials* (New York, John Wiley and Sons), p. 255.
- Guggenheim, E. A., 1957, 3rd edition, *Thermodynamics* (Amsterdam, North-Holland Publishing Co.).

- Haruy, H. K., and Heal, T. J., 1956, in *The Mechanism of Phase Transformations in Metals* (London, The Institute of Metals), pp. 1-46.
- Herring, C., 1952, in *Metal Interfaces* (Cleveland, Ohio, American Society for Metals), pp. 1-19.
- , 1953, in *Structure and Properties of Solid Surfaces* (Chicago, University of Chicago Press), pp. 1-81.
- Hess, J. B., 1952, in *Metal Interfaces* (Cleveland, Ohio, American Society for Metals), pp. 134-152.
- MacDougall, F. H., 1939, 3rd edition, *Thermodynamics and Chemistry* (New York, Wiley and Sons).
- Shaler, A. J., 1953, in *Structure and Properties of Solid Surfaces*, Gomer, R., and Smith, C. S., editors (Chicago, University of Chicago Press), pp. 120-146.
- Smith, C. S., 1952, in *Imperfections in Nearly Perfect Crystals*, Shockley, W., et al., editors (New York, John Wiley and Sons), pp. 377-401.
- Udin, H., 1952, in *Metal Interfaces* (Cleveland, Ohio, American Society for Metals), pp. 114-152.

2.6.2. Original Articles

- Bailey, A. I., 1957, in *Second International Congress on Surface Activity* (London, Butterworths Scientific Publications), Part III, p. 406.
- Cabrera, N., 1959, *Disc. Faraday Soc.*, 28, 1.
- Gilman, J. J., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), pp. 193-222.
- Jura, G., and Garland, C., 1952, *J. Am. Chem. Soc.*, 74, 6033.
- Kozakevitch, P. P., and Urbain, G., 1958, in *The Physical Chemistry of Steel-making*, J. F. Elliot, editor (Technology Press, MIT and John Wiley and Sons, New York), pp. 27-34.
- Oriani, R. A., 1950, *J. Chem. Phys.*, 18, 575.
- Pranatis, A. L., and Pound, G. M., 1955, *Trans. AIME*, 203, 644.
- Riegger, O. K., and Van Vlack, L. H., *Trans. Met. Soc. AIME*, in press.
- Shuttleworth, R., 1949, *Proc. Phys. Soc. (London)*, 62A, 167.
- Skapski, A., 1948, *J. Chem. Phys.*, 16, 387.
- , Billups, R., and Rooney, A., 1957, *J. Chem. Phys.*, 26, 1350.
- Smith, C. S., 1948, *Trans. AIME*, 175, 15.
- Stranski, I. N., 1939, *Ber. deut. chem. Ges.*, A72, 141.

3. SURFACE STRUCTURE

Structural considerations enter into virtually all aspects of surface work, as pointed out in section 1. Here we discuss, in general terms, some of the characteristic structural features of surfaces on the gross or macroscopic as well as the atomic scale. Mention is also made of the techniques currently available for studying the surface topography at both levels of observation.

3.1. Gross Structure

Departures from the ideal atomically smooth surface occur not only on the atomic scale but also on a much larger or coarser scale.

This is true for all surfaces encountered in nature and even for surfaces prepared by the best methods presently available. Thus, the cleavage surfaces of crystals usually exhibit cleavage steps of the order of 10Å to 1000Å or more in height (Tolanski, 1948R; Low 1959R) and of comparable linear dimensions. Annealed surfaces generally exhibit thermal etch patterns and electropolished metal surfaces, even when optically perfect, always show some waviness and may still contain surface protrusions (Thomson, 1953R).

Gross imperfections whose linear dimensions are of the order of 10^{-4} to 10^{-7} cm can be investigated by a variety of different techniques (Bowden and Tabor, 1953R, 1954R). Among these, electron diffraction (Thomson, 1953R), reflection electron microscopy (Menter, 1953), multiple-beam interferometry (Tolanski, 1948R; Bowden, 1956), and low-angle X-ray diffraction (Schulz, 1954), have been widely used. The examination of transparent replicas of the surface by transmission electron microscopy provides another high-resolution tool (sec. 4.5) which is frequently employed (Möllenstedt et al., 1960R), and indirect evidence concerning the surface topography can often be obtained from surface area determinations. For solids which decompose readily in an electron beam, the scanning electron microscope developed by Smith and Oatley (1955) is also a useful instrument (Bowden and McAuslan, 1956).

These various techniques have now reached such a high state of development that it is by no means impossible today to characterize a surface rather completely on a macroscopic level. Preparing surfaces which are free from gross structural imperfections is still, however, exceedingly difficult. Nevertheless, it is probably safe to say that cleavage or electropolishing techniques can, on occasion, produce macroscopic areas which contain almost no or relatively few gross imperfections, although roughness factors of the order of 1.3 are still considered excellent.

3.2. Atomic Structure

The gross structural imperfections which occur in surfaces are more or less accidental and could perhaps be eliminated with suitable refinements in the techniques for preparing surfaces. On the other hand, the detailed atomic configuration at the surface is directly related to the crystal structure itself and is therefore an inherent feature of the particular crystal face exposed at the surface. This is true even in the absence of any deviations from perfect periodicity with respect to the arrangement of atoms

within the crystal. Before discussing what is presently known about the structure of surfaces on an atomic scale, it may be well to indicate briefly the methods currently available for studying the surface structure at this level, as well as their limitations.

The most direct and highest resolution technique yet developed is field ion microscopy (Good and Müller, 1956R; Müller, 1958, 1959, 1960). Under the most favorable circumstances, the resolution obtainable is of the order of 2 or 3 Å so that individual rows of atoms or even single atoms can be resolved on the surfaces of conducting field emitters. Unfortunately the method is presently restricted to moderately strong (i.e., high melting) metals and it also requires small specimens of rather specialized geometry. Another powerful technique is low-energy electron diffraction (Farnsworth, et al., 1958). Its principal usefulness lies in indicating lattice parameters and in revealing imperfections of reasonable number density. While the surface is "seen" much less clearly by low-energy electron diffraction than by the field ion microscope, measurements of the lattice parameters and other resolvable features can be made much more accurately. In addition to direct techniques of this type, indirect methods are also available. Thus, inferences concerning the surface structure can be drawn from work function measurements, from observations on the growth of crystals, and from adsorption and surface diffusion experiments (Gomer, 1961R), as well as from theoretical considerations.

The general conclusions which have been reached on the basis of work carried out to date are the following: deviations from a perfectly ordered atomic arrangement occur on almost all surfaces even if the crystal is free from bulk imperfections and even if precautions are taken to avoid or eliminate adsorbable impurity atoms. These departures from perfect order may consist of extra atoms adsorbed on close-packed planes together with corresponding vacant sites, terraces formed by incomplete planes or rows of atoms, and kinks along the edges of these terraces, which correspond to the ends of incomplete rows of atoms. Defects of this kind are believed to result from thermal disordering of the surface which is "frozen-in" at relatively low temperatures.

Unfortunately, the mobilities of defects are appreciable only at temperatures where considerable disorder already exists. It is therefore impossible to eliminate such defects in finite times by annealing at low temperature. However, it has been demonstrated (Good and Müller, 1956R) that essentially perfect surfaces can be obtained by peeling off one or two lattice planes from a cold surface by means of field desorption. Since this represents a

kinetic process it is not likely, of course, that the resultant surface really corresponds to an equilibrated one.

Thermal disorder can be treated, in principle, by the methods of statistical mechanics. Quantitative calculations are very difficult, although the work of Burton and Cabrera (1949) indicates the general direction in which to proceed. One interesting complication arises from the fact that the mobility of surface atoms can lead to a cooperative surface melting on certain planes at temperatures considerably below the bulk melting point. The molten film can be several atom layers thick. Under these conditions, a quasi-thermodynamic treatment can again be employed (Cahn and Hilliard, 1959).

Deviations from ideality may also arise from other causes. It is possible, for example, that the free energy of the system is minimized, even at low temperature, by changes in the *average* lattice parameter at the surface. This may take the form of a real change in the lattice spacing (although the evidence in favor of this is slim at best) or it may be effected by the introduction of imperfections. At present, little is known about matters of this kind (Cabrera, 1959).

Surface imperfections arise also from those in the bulk. Thus, terraced depressions are produced at the points of emergence of dislocation lines with the surface. Dislocation lines which intersect the surface can be revealed by various techniques (Amelinckx, 1959R) and the local atomic configuration in the neighborhood of dislocations can be studied, in favorable circumstances, by field ion microscopy (Müller, 1958, 1959). Disturbances in the atomic arrangement at the surface may also be associated with the presence of internal crystal boundaries (i.e., grain, subgrain, or twin boundaries) which intersect the surface.

Surface structure is often modified by adsorption or by the presence of epitaxed layers. There is evidence, for example, that oxygen changes the surface structure of tungsten field emitters (Müller, 1959), doubling the step height normally encountered between lattice planes on a spheroidal field emitter. It is possible that this structural change may be brought about by an increase in the area of those particular orientations for which the surface tension is most strongly affected by adsorption.

Marked structural changes also occur during the initial stages of condensation or precipitation on surfaces, as indicated by field emission microscopy (Good and Müller, 1956R). It is clear that the problems encountered with clean surfaces are complicated or magnified to an even greater extent under these conditions. A wide

and largely untapped field for experimental research exists here, since field electron and field ion microscopy are almost ideally suited for attacking problems of this kind.

3.3. Summary

Gross imperfections exist on almost all macroscopic surfaces to a greater or lesser extent. Deviations from ideality on the atomic scale are inherent under almost all working conditions. It is at last becoming possible to obtain clean metal surfaces, at least with sufficiently small specimens such as field emission tips, but there is room for much improvement and for extension of the techniques to a wider variety of materials and to larger specimen sizes. Theoretical predictions of surface disorder are in a highly rudimentary stage, and again much work still remains to be done in order to develop a satisfactory theoretical model.

3.4. References

3.4.1. Reviews and General Articles

- Amelinckx, S., 1959, in *Methods in Experimental Physics*, L. Marton, editor (New York, Academic Press), Vol. 6, pp. 321-356.
- Bowden, F. P., and Tabor, D., 1953, in *Structure and Properties of Solid Surfaces*, R. Gomer and C. S. Smith, editors (Chicago, University of Chicago Press), pp. 203-239.
- Bowden, F. P., and Tabor, D., 1954, *Friction and Lubrication of Solids* (Oxford, Clarendon Press).
- Good, R. H., and Müller, E. W., 1956, *Handbuch der Physik* (Berlin, Springer-Verlag), Vol. XXI, pp. 176-231.
- Gomer R., 1961, *Field Emission and Field Ionization* (Cambridge, Harvard Univ. Press).
- Low, J. R., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), pp. 68-90.
- Möllenstedt, G., Niehrs, H., and Ruska, E., editors, 1960, *Fourth International Conference on Electron Microscopy* (Berlin, Springer-Verlag), Part 1.
- Müller, E. W., 1960, *Adv. Electronics and Electron Physics*, 13, 83.
- Thomson, G. P., 1953, in *Structure and Properties of Solid Surfaces* (Chicago, University of Chicago Press), pp. 185-202.
- Tolanski, S., 1948, *Multiple Beam Interferometry of Surfaces and Films* (Oxford, Clarendon Press).

3.4.2. Original Articles

- Bowden, F. P., 1956, *J. Coll. Sci.*, XI, 555.
- , and McAuslan, J. H., 1956, *Nature*, 173, 408.
- Burton, W. K., and Cabrera, N., 1949, *Disc. Faraday Soc.*, 5, 33.
- Cabrera, N., 1959, *Disc. Faraday Soc.*, 28, 16.
- Cahn, J. W., and Hilliard, J. E., 1959, *J. Chem. Phys.*, 31, 688.
- Farnsworth, H. E., Schlier, R. E., George, T. H., and Burger, R. M., 1958, *J. Appl. Phys.*, 29, 1150.

- Menter, J. W., 1953, *J. Photogr. Sci.*, 1, 12.
Müller, E. W. 1958, *Acta Met.*, 6, 620.
———, 1959, *Z. Physik.*, 156, 399.
Schulz, L. G., 1954, *Trans. AIME*, 200, 1082.
Smith, K. C. A., and Oatley, C. W., 1955, *Brit. J. Appl. Phys.*, 6, 391.

4. THE ROLE OF IMPERFECTIONS IN SURFACE PHENOMENA

The role of surface imperfections in adsorption, in chemical reactivity and in plastic flow and fracture phenomena is considered in this section. It is apparent that theoretical studies of surface imperfections have so far been based upon highly idealized models in which it has been assumed that surfaces are bounded by crystallographic planes of minimum surface energy which are atomically smooth, apart from the presence of parallel surface terraces. In general, it is not possible to produce extended areas of surfaces with such ideal properties using the materials and techniques which are available at the present time. Much of the discussion which follows is therefore concerned with possible future approaches to the experimental problems involved in studying the influence of imperfections on surface phenomena.

4.1. Materials for Crystals

It will not be possible to provide unequivocal evidence for the role of imperfections in surface phenomena until large crystals with habit faces or low energy surfaces of high perfection become available. There are a number of ways in which the surface density of imperfections can then be increased in a deliberate fashion and correlated with any changes in the observed properties. The surface irregularities and high bulk dislocation densities in the majority of the crystals of metals and ionic solids which are available at the present time imply a heavy concentration of surface imperfections which would make it almost impossible to carry out critical experiments in this field.

The ease with which large dislocation-free crystals can be grown and the ease with which dislocation densities can be reduced by subsequent heat treatment (i.e., annealing) depend to a large extent on the nature and concentration of the impurities which are present. The importance of producing materials of high initial purity therefore deserves special emphasis. Methods must also be developed for minimizing contamination with impurities from substrates, crucibles, furnace linings, gas atmospheres and related sources during crystal growth, shaping, surface finishing and annealing.

The materials which appear to be most suitable for experimental work on surface imperfections at the present time are: LiF (Gilman and Johnston, 1957) and NaCl (Amelinckx, 1954, 1956); AgCl and AgBr (Hedges and Mitchell, 1953a, 1953b; Mitchell, 1957, 1958a); Cu, Ag, Au and Al (Forty and Frank, 1953; Hirsch, 1959; Lovell and Wernick, 1959; Lang, 1959; Young and Gwathmey, 1960); Fe, W and Mo (Low and Guard, 1959; Schadler, 1960); Te (Davies, 1957; Stokes, Johnston and Li, in press); and Si and Ge (Vogel, 1955; Tyler and Dash, 1957; Dash, 1958, 1959; Bennett and Longini, 1959). These substances crystallize in a variety of different structures and single crystal specimens with relatively low dislocation densities, grown from high-purity starting materials, should be obtainable for experimental work designed to establish the nature and role of imperfections in surface phenomena. Other materials can be added successively to this list as the necessary effort is put into their purification and their production in the form of single crystals.

4.2. Growth of Crystals

For most experimental work on surface imperfections, it would be desirable to use habit faces produced naturally during the growth of crystals from the vapor. Such surfaces require no further mechanical or chemical treatments, which are likely to introduce high densities of dislocations and impurities, before they are used for experimental studies. Crystals of Ag (Forty and Frank, 1953), Mg (Forty, 1952), Zn and Cd (Coleman and Sears, 1957; Cabrera and Price, 1958) can be grown from the vapor in the form of single crystal platelets with relatively large flat surfaces bounded by close-packed planes of atoms. Iron may be produced in the form of thick single crystal needles with surfaces of high perfection by the reduction of iron halides (Coleman, 1958; Laukonis and Coleman, 1959). In both cases, the crystals are grown upon a substrate of the material itself; as has been shown clearly with Si, this minimizes contamination. Some further purification of the initial materials is, moreover, probably effected by fractional distillation or preferential reduction when crystals are grown by techniques of this kind. Recently, crystals of germanium in the form of dendrites having atomically flat surfaces have been grown from the melt (Bennett and Longini, 1959).

For the majority of materials, however, surfaces for study must currently be produced by sectioning large single crystals. Thus, crystals of the alkali halides may be grown from the melt by the methods of Bridgman and Stockbarger or of Czochralski and

Kyropoulos (Tanenbaum, 1959) and then annealed, cleaved to produce plane surfaces, and re-annealed to remove strains and dislocations introduced during cleavage. Crystals of silver halides may also be grown by these methods, but since these materials cannot be cleaved, surfaces must be produced by a combination of mechanical and chemical methods. It has been found more profitable to grow thin sheet crystals having optically flat surfaces by crystallizing discs of the molten silver halides between glass plates (Clark and Mitchell, 1956).

Small crystals of metals may be grown by electrodeposition, by condensation from the vapor, or by chemical reactions occurring at the interface between the crystal and the surrounding medium (Brenner, 1956). Large metal single crystals grown from the melt are usually found to contain a high density of dislocations, due either to the initial impurity content, to contamination with crucible constituents during growth, to the internal condensation of vacancies which gives rise to the formation of dislocation loops, or to the relaxation of thermal strains arising from non-uniform cooling and from differential contraction between crucible or substrate and the crystal during cooling. Much work remains to be done before crystals of the majority of the metals become available which will be suitable for the study of the role of imperfections in surface phenomena.

The dislocation density of many materials may be reduced by annealing heavily deformed specimens at the lowest temperature at which recrystallization *in situ* will proceed. At these relatively low temperatures, the concentration of vacant lattice sites is many orders of magnitude less than that at the melting point; hence, the density of dislocation loops formed by the subsequent condensation of thermal vacancies is correspondingly smaller. This strain-anneal method has been used successfully for producing crystals of AgCl and AgBr with very low densities of dislocations (Mitchell, 1958a).

The control of chemical purity and structural perfection during crystal growth is discussed at greater length in Part VII dealing with the "Growth, Structure and Morphology of Crystals."

4.3. Production of Surfaces

When naturally formed habit faces are not available and the crystals cannot be cleaved, surfaces parallel to crystallographic planes of low energy must be produced by mechanical means such as sawing, milling, and abrasion. These operations generally cause heavy surface damage and lead to the introduction of high

densities of dislocations in all materials which deform plastically at room temperature. The disturbed surface layers must therefore be removed by polish-etching with suitable polishing laps and solutions, followed by chemical or electrolytic polishing. Surfaces approximately parallel to certain crystallographic planes, representing the energetically stable interfaces under the conditions of attack or dissolution, may then be developed either by suitable etching methods or by reactions involving active gases which produce volatile or non-adherent products. Surfaces which appear to be parallel to crystallographic planes may also be produced on single crystals of a number of metals by electrodeposition. The surfaces produced by electrolytic polishing or electrodeposition, and the apparently crystallographic surfaces produced by chemical attack, are usually covered with adsorbed films of impurities which must be removed by suitable chemical methods before the crystals are annealed.

The final stage in the production of a stable surface, which may be expected to provide a reasonable approximation to the theoretical model, usually consists of a long period of annealing in high vacuum at the minimum temperature at which the surface atoms are freely mobile, followed by cooling at a very slow rate. As pointed out in section 3.2, it is also possible to obtain essentially perfect surfaces by field desorption, at least with micro-specimens of high melting metals.

4.4. The Nature of the Surface Imperfections

The imperfections on habit faces and surfaces of low energy which are of greatest importance are terraces formed by incomplete planes of atoms or ions, kink sites along the edges of these terraces, which correspond to the ends of incomplete rows of atoms or ions, and terraced depressions associated with the points of intersection of dislocation lines with the surface (Burton, Cabrera, and Frank, 1951). The density of terraces and kink sites must be greater on the curved surfaces of these depressions than elsewhere on the surfaces of the crystals, and it is this property which is probably mainly responsible for any role which dislocations may have in surface chemical reactivity (Forty, 1954R; Evans and Mitchell, 1955; Frank, 1955R; Mitchell, 1955R).

4.5. Observations of Surface Imperfections

Local curvatures and etch pits can be observed on the surfaces of large specimens with the optical microscope using the methods of phase contrast microscopy, interference microscopy and mul-

multiple-beam interferometry, but these methods are not sufficiently sensitive to permit the observation of monolayer terraces on the surfaces of metals and simple ionic crystals. The interference microscope has many advantages as a non-destructive tool for the examination of relatively coarse surface structures; for example, since no preliminary treatment of the surface is required, the original surface may be used for further experimental work.

Monatomic terraces on crystals of silver halides can be observed with the optical microscope only after they have been decorated with silver; such terraces on habit faces of metallic crystals may be rendered visible by localized chemical attack or by the deposition of some decorating substance along their edges (Forty and Frank, 1953; Forty, 1952). Monatomic terraces have also been observed on the cleavage surfaces of sodium chloride crystals by first depositing gold atoms along the terraces and then examining carbon replicas of the surface in a high-resolution electron microscope (Bassett, 1958). Similar methods could no doubt be found for decorating terraces on the surface of other crystals but all these methods modify the composition and structure of the surface and make it unsuitable for further experimental work.

The high-resolution electron microscope has also been used to make direct observations of surface terraces and associated dislocations (Hirsch, 1959R). The method used does not necessitate any decoration of the surface terraces but the specimens must be thinned from one side by electrolytic polishing while the other side is protected from attack by the polishing solution (Thomas and Hale, 1959; Hirsch et al., 1959); again, the resulting specimen may not be suitable for further experimental study. It has, however, been possible to observe surface terraces in carefully thinned specimens under favorable conditions of dark-field illumination.

The recent development of the Berg-Barrett technique by Newkirk (1958, 1959) has provided a powerful non-destructive X-ray method for studying dislocations in the subsurface regions of crystals which does not require any thinning of the specimens. It produces a photographic record at unit magnification but cannot be generally applied at present because the dislocation densities in metallic crystals are usually too high. Similar X-ray methods developed by Lang (1958, 1959), Bonse and Kappler (1958), Borrmann et al. (1958) and Barth and Hosemann (1958) also allow reliable estimates to be made of the dislocation densities and configurations in large specimens with low dislocation densities. These methods are certain to be used to an ever-increasing extent in the future, particularly as the dislocation densities of crystals

are progressively reduced by improvements or refinements in the methods used for the purification of the initial materials and in the techniques for growing, shaping, finishing and annealing large single crystals.

With some metals, field emission and field ion microscopy (Müller, 1958, 1960) can also be used to study surface imperfections. Specimens of very small size are required, however, and as a consequence, the sample may contain only a very small number of dislocations, or perhaps none at all. However, it is possible to introduce imperfections by reverse field stressing (sec. 8.5).

4.6. Introduction of Controlled Densities of Imperfections

The establishment of the role of imperfections and dislocations in surface phenomena will have to await the production of crystals of higher purity and lower dislocation density than those which are available at present. When suitable crystals become available, it should be possible to produce surfaces which approximate closely the idealized models used in the theoretical investigations. It should then be possible to introduce controlled densities of dislocations and other surface imperfections and to make experimental observations which could perhaps be interpreted directly and unambiguously in terms of the theoretical models.

As an example of a possible experimental approach, we may consider a single crystal of a face-centered cubic metal. A specimen in the form of a rectangular parallelepiped may be cut from the crystal with upper and lower surfaces parallel to (111) and ($\bar{1}\bar{1}\bar{1}$) planes, lateral surfaces parallel to ($\bar{1}10$) and ($1\bar{1}0$) planes, and end surfaces parallel to ($\bar{1}\bar{1}2$) and ($11\bar{2}$) planes, respectively. If this single crystal specimen were prepared as previously outlined and annealed for a long period in ultra-high vacuum, in the absence of active gases which could be adsorbed on the surfaces, it might be anticipated that the (111) surfaces would break up into a limited number of terraces bounded by [$\bar{1}10$], [$0\bar{1}1$] and [$\bar{1}01$] directions. If this situation were realized in practice, the density of terraces parallel to the [$\bar{1}10$] direction could then be increased in systematic fashion by preparing a series of specimens whose upper surfaces make gradually increasing angles with the (111) plane (due to rotation about the [$\bar{1}10$] direction lying in the plane). The actual step heights would depend upon the occurrence of bunching during the annealing period, and the surface contours would have to be established with the aid of the electron microscope, using either a replica technique or some appropriate thinning procedure. This work could be paralleled by a

theoretical study of the equilibrium form of the surface of a face-centered cubic crystal as a function of the angle between the normal to the surface and the $[111]$ direction.

Controlled densities of dislocation loops could be introduced at the surfaces of the essentially dislocation-free specimens by uniform bending about an axis parallel to the $[\bar{1}10]$ direction. The dislocation loops so formed could be observed directly by the methods of Newkirk, Lang and others, or by examination with the electron microscope after the specimen has been suitably thinned. The surface reactivity of similarly prepared specimens could then be studied and compared with that of dislocation-free specimens.

By developing such methods of investigation, of which these are given only as suggestive examples, it should be possible to produce metal specimens with which the role of surface imperfections and dislocations in surface phenomena could be established experimentally.

4.7. The Role of Imperfections

The reaction of molecules in the vapor phase or of molecules or ions in solution with the surface of a crystal involves the adsorption of the reactants at the surface and the formation and desorption of the reaction products. The importance of surface terraces, with their associated kink sites, in determining the kinetics of the reaction must depend on the temperature at which the reaction occurs, on the heat of adsorption of the reactants at the surface, on the heat of reaction and on the heats of desorption of the products. Many different situations can be envisaged.

Well-defined steps or terraces lying along close-packed directions and having kink sites at regular intervals should be present upon the surfaces of dislocation-free crystals which have been annealed in vacuum for long periods and then cooled slowly to temperatures at which atoms do not have sufficient energy or mobility to leave the edges of the terraces and migrate away. Surface mobility will increase when the temperature is raised, and the concept of a surface step as a linear imperfection parallel to a close-packed row of atoms may then lose its significance. The edge of the terrace will be very irregular but there will always be a definite density of kink sites along the edge of the terrace under any particular group of physicochemical conditions.

The heats of adsorption of the reacting atoms or molecules depend upon the integrated residual combining capacity of the atoms neighboring each adsorption site. On this basis, the various

sites may be arranged in order of decreasing heats of adsorption; sites adjacent to kinks have the largest, sites along the edges of terraces somewhat smaller, and sites on close-packed surfaces still smaller heats of adsorption. Molecules will therefore be adsorbed preferentially at kink sites when the temperature and the partial vapor pressure or the concentration of the reactive species are such that the coverage of the surface is small and dynamic equilibrium is maintained with the vapor phase. The energy required for the detachment of reacting atoms from the surface of the crystals is less for atoms at kink sites and at sites along the edges of surface terraces than for other surface sites; consequently, reaction should occur preferentially at kink sites and along the edges of surface terraces when the surface coverage is small or when a complete monolayer is present. Reaction under these conditions would result initially in the decoration of the surface terraces with reaction products and finally in the formation of monolayers of reaction products on the surfaces. If the reaction products were then either volatilized or detached from the surface, or if the reacting molecules could diffuse freely through the product layer, reaction would proceed through the combination of the surface atoms with the reacting molecules row-by-row and layer-by-layer.

The experience which we have gained in developing reliable methods for the decoration of monatomic surface terraces and dislocations with reaction products seems to indicate that the idealized conditions outlined above can seldom be realized in practice. In the majority of the systems studied, reaction either occurs uniformly over the whole surface or small islands of reaction product are formed near the edges of the surface terraces, thus rendering the terraces visible. That there should, in general, be no localized reaction is perhaps not surprising. In any exothermic reaction, the formation of a monolayer of adsorbed molecules on an atomically smooth surface will probably be followed by reaction with atoms in the underlying surface layer, so that an island of reaction product is formed in a depression one atom layer deep which is bounded by surface terraces. This island would then increase in area as the reaction progresses; the surface terraces involved would be introduced by the reaction itself and pre-existing terraces would play at most only a minor role.

On the other hand, surface terraces do appear to play a significant role in determining the kinetics of reactions under some physicochemical conditions or when the reaction itself is of an autocatalytic nature. In such circumstances, nuclei may be formed

along the surface terraces and grow to visible size as the reaction proceeds, and in this way surface steps, together with the points of intersection of dislocations with the surface, are made visible. Depressions should be produced at these points when the crystals are annealed in a good vacuum at temperatures at which the surface atoms are freely mobile. The local surface density of steps should be greater on the curved surfaces of the depressions than elsewhere, so that the reaction products should be concentrated at the dislocations. Successive, outward-spreading terraces or continuously expanding spiral terraces, which separate reacted from unreacted surface areas, would be associated with dislocation lines emergent at the surface, the type of terrace depending on the orientation of the Burgers vector of the dislocation with respect to the surface of the crystal. In these circumstances then, the rate of formation of nuclei and the initial rate of reaction would be functions of the density and configuration of dislocations at the surface.

The validity of any such predictions concerning the role of surface terraces could possibly be established by experiments with dislocation-free crystals having surfaces which make small angles with stable crystallographic planes, as discussed in section 4.6. The surface density of nuclei and the extent of reaction after a definite time under particular physicochemical conditions should then be functions of the angle between the normals of the surface and to the stable plane. The role of dislocations could also be investigated using low dislocation density crystals which had been bent so as to introduce known numbers of dislocations in known configurations.

In certain circumstances, a reactant may produce nuclei of the new phase along dislocation channels below the surface as well as on the free surface of the crystal. The product would strain the lattice if it occupied a greater volume than the atoms of the substrate from which it was formed. This would increase the free energy of the crystal in the neighborhood of the dislocation lines; a further increase in the probability for the initiation of surface reaction terraces might be produced as a consequence.

There are other important physical phenomena which may depend upon or be influenced by the chemical reactivity of surface imperfections. For example, the available experimental evidence seems to indicate that the plastic deformation of a dislocation-free crystal is initiated with the formation of dislocation loops on glide planes at or near the surface (Gilman and Johnston, 1957; Johnston and Gilman, 1960). The generation of many of these loops is

accompanied by the appearance of a step or surface terrace associated with the component of the Burgers vector of the dislocation line normal to the surface. It would be of the greatest possible importance to establish whether or not the generation of such dislocation loops is influenced by the composition of the atmosphere in which the crystals are strained. Crystals prepared as discussed in section 4.6 could be used for these experiments. The adsorption of reactive gas molecules along the edges of surface terraces associated with the formation of very small dislocation loops might lower the critical stress required for the continued expansion of the loops and consequently reduce the yield stress of the material. Experiments such as these, which are designed to establish the connection between the surface density of terraces, the partial pressure of adsorbable gases and the yield stress, should provide results of considerable value and would undoubtedly contribute much to our understanding of adsorption, of surface chemical reactivity, of fatigue and stress-corrosion, and of other yield point phenomena during plastic deformation.

4.8. Silver Halide Crystals as a Model System and General Conclusions

Many properties of monatomic surface steps and dislocations have been established by experiments carried out with thin sheet crystals of AgCl and AgBr having very low dislocation densities which have been prepared by the strain-anneal method. Dislocation loops are produced at the surfaces when the crystals are deformed by the application of a tensile stress. The monatomic terraces associated with the loops, as well as the loops themselves, are made visible by the photolytic separation of particles of silver when the crystals are exposed to light (Bartlett and Mitchell, in press).

The immersion of the crystals for a fraction of a second in a solution of $\text{Na}_2\text{S}_2\text{O}_3$ causes localized dissolution around the points of intersection of the dislocations with the surface; shallow etch pits are thus formed which may be studied with the interference and phase contrast microscopes. Pits are not produced on the surfaces of dislocation-free crystals. The number of pits increases continuously when the crystals are plastically strained because deformation results in the generation of dislocation loops at the surfaces (Parasnis and Mitchell, 1959).

Well-defined, localized etch pits are produced because the outward spreading of the surface steps is inhibited by the adsorption of molecules of $\text{Ag}_2\text{S}_2\text{O}_3$ along the steps. When the crystals are

washed, the $\text{Ag}_2\text{S}_2\text{O}_3$ molecules must break down to form adsorbed molecules of Ag_2S , because a surface latent image which will initiate chemical development is now formed when the crystals are exposed to light. There is a higher probability of forming a surface latent image at points where dislocations meet the surface than elsewhere, this being consistent with the decoration of the surface steps with adsorbed molecules of silver sulphide (Mitchell, 1957R, 1958b). An internal latent image is also formed during exposure along the dislocation lines in the subsurface regions of the crystals (Hedges and Mitchell, 1953a, 1953b).

In summary, the observations which have been made on the decoration of monatomic surface steps and on the role of dislocations in determining the chemical and the photochemical reactivity of large single crystals of silver halides leave little doubt as to the role which steps and dislocations might play in surface phenomena in other crystals. The silver halide crystals do, however, provide a very sensitive system for studying such phenomena and it would be quite premature to generalize the conclusions which have been reached on the basis of these studies or to assume, without experimental confirmation, that they could be extended to a wide range of chemical reactions occurring on the surfaces of other substances. Such a generalization would be immediately confronted with many exceptions and this would only lead to skepticism about its validity and utility.

More experimental evidence must therefore be obtained before the role of the imperfections can be clearly defined, and the main initial effort must be concentrated upon the purification of materials and the production of crystals with very low dislocation densities. Carefully planned experiments should then yield conclusive results and provide a sound foundation for theoretical investigations on the structure of surfaces. It would be desirable that the experimental work be planned in such a way that the results have a direct bearing on the theoretical studies because it is only when there is strong interaction between these two approaches that rapid progress is usually made. A carefully coordinated program of this kind can scarcely fail to be profitable in the long run.

4.9. Structure-Dependent Chemical Activity of Metal Surfaces

Imperfections within the bulk or volume of the material will intersect the surface along a point or a line. The surface may also contain imperfections lying in its plane, e.g., the steps of a growth spiral. It has been indicated in general terms why these regions

may be the sites of intensified reactivity with liquid or gaseous environments. We shall now consider more specifically the reactivity in the vicinity of structural imperfections in otherwise homogeneous metals and alloys.

Beginning with an arbitrary plane surface, two general cases can be distinguished: (a) The reaction proceeds rapidly in the vicinity of the imperfection, but the initially rapid rate gradually diminishes and finally becomes equal to that on adjacent areas of the crystal surface; (b) the rapid rate of reaction at the imperfection, relative to the adjacent structure, is maintained until the solid disintegrates. The first situation represents the attainment of a steady-state surface configuration, like a thermally etched groove at a grain boundary, and does not seriously affect the mechanical stability; the second corresponds to a specific instability of the solid in the vicinity of the imperfections, and can result in fracture or rapid disintegration.

At the present time, it is not possible to predict in detail how any given material would behave when exposed to a specific environment on the basis of known chemical and structural principles. Further, since localized reactivity (penetration per unit time) is dependent on stress, the ordinary concept of static strength ceases to have any useful meaning in this connection.

In many applications of materials, the problem of microstructural stability has become a limiting factor in design and application; for example, an incomplete list of troublesome phenomena would include transgranular fracture of stainless steel, total and rapid disintegration of pure aluminum in water at elevated temperatures, penetration of solids by liquid metals in reactors, delayed fracture of glass and plastics, corrosion fatigue and intergranular corrosion. In all cases failure can be associated with localized chemical reaction; moreover, virtually all types of material seem to be susceptible in some environment. Thus, the outlook is gloomy unless the factors controlling these processes are brought into the class of predictable events by investigations designed to fill in the existing gaps in our knowledge of the structure-dependent chemical reactivity of solids.

4.9.1. Current Knowledge

The surface reactivity of a pure monatomic substance, typically a metal, seems to be determined, at least, in part by the crystallography of the surface. That is, the reactivity in the vicinity of imperfections, whether isolated dislocations or arrays of dislocations at a grain boundary or subboundary, depends critically

on the orientation of the surface; when unstable configurations are eliminated, e.g., by dissolution, the specific reactivity of the site often disappears (Bakish and Robertson, 1956; Tragert and Robertson, 1955). In this case, more or less flat-bottomed etch pits are produced, and prolonged exposure to the chemical environment does not cause penetration at the site of the imperfection.

In an alloy or impure metal, however, the reactivity in the vicinity of the imperfection may remain at a high level, relative to the surrounding matrix. Thus, it appears that pure metals are not subject to the same processes of disintegration as alloys. In a given environment, localized reactivity of a solid solution alloy depends on the composition of the alloy and the type of solute which is present; for example, 0.002% P is sufficient to cause intergranular corrosion of a Cu-P alloy in ammonia, whereas about 5% Zn is required to produce similar susceptibility in a Cu-Zn alloy (Thompson and Tracy, 1949).

Further investigation is required to clarify such questions as:

(1) The mechanism of selective attack at a grain boundary in a homogeneous solid solution that produces a high rate of penetration relative to the adjacent crystals.

(2) The structural as well as chemical reasons for the composition dependence of localized reactivity in alloys.

(3) The possible role of composition gradients at grain boundaries (adsorption of alloy components).

(4) The effect of a third component added to a binary solid solution, which often results in greatly reduced reactivity at a grain boundary.

Plastic deformation of a homogeneous solid solution increases the number of dislocations and also introduces other types of imperfections, e.g., point defects such as vacancies and interstitials, stacking faults, deformation bands, slip bands and arrays of dislocations piled up at internal barriers. Since characterization of the structural nature of the cold-worked state is difficult, the reactivity of such aggregates of imperfections has not been investigated in any detail and their role or function in the fracture process is largely unknown.

Plastic deformation may produce transgranular fracture (Althof, 1944; Robertson, unpublished) in an alloy which fails by intergranular corrosion in the absence of deformation. It appears that deformation creates new sites which are even more susceptible to localized attack than the grain boundaries in stress-free material. Alternatively, brittle fracture may be produced by a chemical reaction only within the grains themselves, or in a single

crystal, in which case the relationship between the nucleating site for fracture and the structural state of the strained crystal requires clarification; for example, is fracture nucleated at sites where dislocations emerge at the surface, or does fracture propagate from the surface, where dislocation arrays may be blocked, producing a high level of reactivity and a rapidly growing crack? In some alloys, extensive deformation may greatly diminish susceptibility to either intergranular or transgranular failure. Thus, depending on the alloy, plastic deformation may either change the mode of fracture or entirely suppress the susceptibility to stress-corrosion.

This entire field seems to be plagued by empirical observations which are often conflicting or contradictory in nature. The vast majority of experiments reported to date have been carried out on materials of poorly-characterized composition and structural perfection or under poorly defined or controlled physicochemical conditions. Systematic investigations appear to be almost entirely lacking and a great deal of fundamental information must first be accumulated before it will be possible to sort out the significant details from the large body of empirical observations reported in the literature. The complexity of the problem will then require that new experiments be designed specifically to verify any apparent or suggested mechanisms for localized chemical reactivity associated with structural imperfections in metals and alloys.

4.9.2. Possible Areas for Future Investigation

With the techniques which have recently been developed for controlling the chemical purity, composition, and structural perfection of crystals and for characterizing the structural state of the material, it should now be possible to carry out meaningful experiments in this area. For example, systematic studies of the reactivity of well-defined arrays of imperfections, such as dislocation boundaries, in both pure metals and homogeneous solid solution alloys could readily be undertaken. Single crystals could be employed for this purpose and dislocation boundaries of known misorientation and structure could be introduced by appropriate mechanical and thermal treatments (Li, Washburn and Parker, 1953). Moreover, bicrystals (or multicrystals) can be grown in which each crystal has a prescribed geometry and controlled orientation both with respect to the component or neighboring crystals and any externally imposed stress axis (Elbaum, 1960; Wagner and Chalmers, 1960). Specimens of this type would be ideal for studying the susceptibility of grain boundaries to cor-

rosion in pure metals and alloys, both in the presence or absence of an externally applied stress. Effects associated with the segregation of impurities at grain boundaries could be investigated in this manner and the possible influence of interactions between point defects (i.e., vacancies) and dislocations could similarly be examined (Arora and Metzger, 1960).

Experiments of this type are merely indicative of the kind of approach which will be required in the future. Since many corrosion processes in alloy systems appear to be associated with the selective removal of one of the components of the alloy, it will also be necessary to investigate in detail the mechanism and kinetics of such reactions. In particular, information is needed concerning the detailed atomic mechanism by which one component is selectively removed from solid solution as well as the structural state of the remaining material.

Continued reactivity at an imperfection is evidently dependent upon composition or upon the nature and distribution of impurities which are present. At the present time, however, relatively little is known with certainty about the interactions between impurities or solute atoms and structural imperfections of various kinds. Information is badly needed concerning the segregation of impurities at grain boundaries in metals, and about grain boundary compositions and energies in alloy systems as a function of the boundary structure or orientation. Further study of the nature of solute atmospheres around isolated dislocations or dislocation arrays is also merited. In this connection, quenching experiments in alloy systems may prove to be enlightening.

At the outset, fracture processes associated with the reactivity of imperfections introduced by plastic deformation require careful study in order to distinguish between cause and effect. We may well ask, for example, whether brittle fracture under a combined state of stress and chemical environment occurs by the propagation of a crack which is produced at the surface by localized reactivity at some structural imperfection, or whether fracture results from increased reactivity associated with the high state of strain at the leading edge of a crack that is produced by a purely mechanical process. In other words, in a plastically deformed material or in a material under stress, does the presence of the chemical environment exert a major influence on the nucleation or initiation of the fracture crack or upon the rate of propagation of the crack, or both?

Finally, the phenomena of interest here are not limited or confined to metals. Since fracture processes are somewhat better

understood in alkali halides and other brittle solids (Averbach et al., 1959R), studies dealing with the relation between surface reactivity and fracture behavior of such materials might well be of considerable value (Stokes, Johnston and Li, 1960). Ionic solid solutions could also be employed as analogues of metallic solid solutions.

4.10. References

4.10.1. Reviews and General Articles

- Averbach, B. L., Felbeck, D. K., Hahn, G. T., and Thomas, D. A., editors, 1959, *Fracture* (Technology Press, MIT and John Wiley and Sons, New York)
 Forty, A. J., 1954, *Adv. in Phys.* **2**, 1.
 Frank, F. C., 1955, in *Chemistry of the Solid State* (London, Butterworth's Scientific Publications), p. 1.
 Hirsch, P. B., 1959, *J. Inst. Metals*, **87**, 406.
 Mitchell, J. W., 1955, in *Chemistry of the Solid State* (London, Butterworth's Scientific Publications), p. 311.
 ———, 1957, in *Die Photographische Empfindlichkeit* (Darmstadt, Helwich).
 Müller, E. W., 1960, *Adv. in Electronics and Electron Physics* **13**, 83.
 Tanenbaum, M., 1959, in *Methods in Experimental Physics*, L. Marton, editor (New York, Academic Press), Vol. 6, Part A, pp. 86–121.

4.10.2. Original Articles

- Althof, F. C., 1944, *Z. Metallkunde*, **36**, 177.
 Amelinckx, S., 1954, *Acta Met.*, **2**, 848.
 ———, 1956, *Phil. Mag.*, **1**, 269.
 Arora, O. P., and Metzger, M., 1960, *Acta Met.*, **8**, 49.
 Bakish, R., and Robertson, W. D., 1956, *J. Electrochem. Soc.*, **103**, 320.
 Barth, H., and Hosemann, R., 1958, *Z. Naturforsch.*, **13a**, 792.
 Bartlett, J. T., and Mitchell, J. W., in press.
 Bassett, G. A., 1958, *Phil. Mag.*, **3**, 1042.
 Bennett, A. L., and Longini, R. L., 1959, *Phys. Rev.*, **116**, 53.
 Bonse, U., and Kappler, E., 1958, *Z. Naturforsch.*, **13a**, 348.
 Borrmann, G., Hartwig, W., and Irmeler, H., 1958, *Z. Naturforsch.*, **13a**, 423.
 Brenner, S. S., 1956, *Acta Met.*, **4**, 62.
 Burton, W. K., Cabrera, N., and Frank, F. C., 1951, *Phil. Trans. Roy. Soc.*, **A243**, 299.
 Cabrera, N., and Price, P. B., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 204.
 Clark, P. V., and Mitchell, J. W., 1956, *J. Photogr. Sci.*, **4**, 1.
 Coleman, R. V., 1958, *J. Appl. Phys.*, **29**, 1487.
 ———, and Sears, G. W., 1957, *Acta Met.*, **5**, 131.
 Dash, W. C., 1958, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 361.
 ———, 1959, *J. Appl. Phys.*, **30**, 459.
 Davies, T. J., 1957, *J. Appl. Phys.*, **28**, 1217.
 Elbaum, C., 1960, *Trans. Met. Soc. AIME*, **218**, 444.
 Evans, T., and Mitchell, J. W., 1955, in *Defects in Crystalline Solids* (London, The Physical Society), p. 409.

- Forty, A. J., 1952, *Phil. Mag.*, **43**, 949.
- , and Frank, F. C., 1953, *Proc. Roy. Soc.*, **A217**, 262.
- Gilman, J. J., and Johnston, W. G., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 386.
- Hedges, J. M., and Mitchell, J. W., 1953a, *Phil. Mag.*, **44**, 223; 1953b, *ibid.*, **44**, 357.
- Hirsch, P. B., Partridge, P., and Segall, R., 1959, *Phil. Mag.*, **4**, 721.
- Johnston, W. G., and Gilman, J. J., 1960, *J. Appl. Phys.*, **31**, 632.
- Lang, A. R., 1958, *J. Appl. Phys.*, **29**, 597.
- , 1959, *J. Appl. Phys.*, **30**, 1748.
- Laukonis, J. V., and Coleman, R. V., 1959, *J. Appl. Phys.*, **30**, 1364.
- Li, C. H., Washburn, J., and Parker, E. R., 1953, *Trans. AIME*, **197**, 1223.
- Lovell, L. C., and Wernick, J. H., 1959, *J. Appl. Phys.*, **30**, 590.
- Low, J. R., and Guard, R. W., 1959, *Acta Met.*, **7**, 171.
- Mitchell, J. W., 1957, in *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., editors (New York, John Wiley and Sons), p. 69.
- Mitchell, J. W., 1958a, in *Growth and Perfection of Crystals* (New York, John Wiley and Sons), p. 386.
- , 1958b, *J. Photogr. Sci.*, **6**, 57.
- Müller, E. W., 1958, *Acta Met.*, **6**, 620.
- , 1959, *Z. Physik.*, **156**, 399.
- Newkirk, J. B., 1958, *Phys. Rev.*, **110**, 1465.
- , 1959, *Trans. Met. Soc. AIME*, **215**, 483.
- Parasnis, A. S., and Mitchell, J. W., 1959, *Phil. Mag.*, **4**, 171.
- Robertson, W. D., unpublished.
- Schadler, H. W., 1960, *Trans. Met. Soc. AIME*, **218**, 649.
- Stokes, R. J., Johnston, T. L., and Li, C. H., 1960, *Trans. Met. Soc. AIME*, **218**, 655.
- Stokes, R. J., Johnston, T. L., and Li, C. H., in press.
- Thomas, K., and Hale, K. F., 1959, *Phil. Mag.*, **4**, 531.
- Thompson, D. H., and Tracy, A. W., 1949, *Trans. AIME*, **185**, 100.
- Tragert, W. E., and Robertson, W. D., 1955, *J. Electrochem. Soc.*, **102**, 86.
- Tyler, W. W., and Dash, W. C., 1957, *J. Appl. Phys.*, **28**, 1221.
- Vogel, F. L., 1955, *Acta Met.*, **3**, 95.
- Wagner, R. S., and Chalmers, B., 1960, *J. Appl. Phys.*, **31**, 581.
- Young, F. W., and Gwathmey, A. T., 1960, *J. Appl. Phys.*, **31**, 225.

5. STRUCTURE AND PROPERTIES OF GRAIN BOUNDARIES

Thus far, the discussion has dealt largely with crystal-fluid interfaces. There are, however, several types of crystal-crystal interfaces and, as noted in the preceding section, these often play a vital role in many reactions or processes which occur on or in solids, e.g., corrosion and oxidation processes, precipitation reactions and other phase transformations, recrystallization and grain growth, plastic deformation and fracture, etc. The only type of crystal-crystal boundary which has so far been studied in any detail is that separating two crystals of the same composition and

structure, but of different orientation, i.e., grain boundaries. We will now consider what is presently known about grain boundaries, concentrating our attention on the structure and properties of such boundaries rather than on the influence which grain boundaries exert on bulk properties per se.

The discussion will be limited to grain boundaries in metals and in germanium, as dictated largely by the scarcity of information available for other classes of materials such as the alkali or silver halides, oxides, etc. Interphase boundaries, i.e., boundaries separating crystals of different crystal structure (as well as composition and orientation), have also been neglected, not because they are any less important but because so little is actually known about their structure and properties. It is hoped, however, that the following summary and evaluation of the existing knowledge concerning grain boundaries might perhaps stimulate attempts to study interphase boundaries in greater detail.

A comprehensive review of the literature pertaining to grain boundaries has recently been published by Amelinckx and Dekeyser (1959R). An extended discussion of the factors which influence the migration or movement of crystal-crystal boundaries may also be found in Part VII, "Growth, Structure and Morphology of Crystals," section 4.4, of this volume.

5.1. Structure of Grain Boundaries

5.1.1. Low-Angle Grain Boundaries

When the crystal lattices on opposite sides of a grain boundary differ in orientation by not more than a few degrees, the boundary separating the two crystals is classified as a "low-angle" boundary. That boundaries of this type can be described geometrically by arrays of dislocations has been known for some time. Although experimental support has come more slowly, there is now no reasonable doubt that the low-angle boundaries observed in solids do, in fact, consist of dislocation arrays. The possible ways of making up such boundaries with dislocations have been discussed at some length by Read (1953R) and by Amelinckx and Dekeyser (1959R).

The more important types of experiments which have helped to establish the dislocation model are the following:

(a) Regularly spaced etch pits have been observed to form along certain types of low-angle boundaries, with a spacing corresponding to that calculated on the assumption that each pit represents an individual dislocation (Vogel, Pfann, Corey and Thomas, 1953).

(b) The movement of certain types of boundaries under the action of an applied stress (Li, Edwards, Washburn and Parker, 1953) is consistent with a description of the boundary in terms of dislocation arrays.

(c) The energy of the boundaries can be accounted for on the assumption that they consist of dislocations (Gjostein, 1958; Gjostein and Rhines, 1959; Wagner, 1959; Wagner and Chalmers, 1960). Etch pit experiments have given excellent quantitative agreement with the theoretical model. Qualitative agreement is also provided by experiments on the motion of dislocation boundaries under stress, but the detailed mechanism of the migration in real crystals is apparently more complex than that predicted by the simple model (Li, 1960).

The calculation of the energy of a dislocation boundary as a function of misorientation is more difficult. The best-known calculation is that of Read and Shockley (1950, 1952R). Van der Merwe (1950) has given a more general treatment which reduces to the Read-Shockley equation for very small misorientations and extends the range of validity to larger misorientations. It was not until quite recently, however, that measurements were made at small enough misorientations to check the predictions in the range where the theory should be applicable.

Several papers and reviews were published in the early 1950's which indicated that the experimental data could be fitted by the Read-Shockley equation up to misorientations as large as 20 to 30 degrees.¹ However, it should be emphasized that the early work provided only relative measurements of the surface tension γ , so that there were two disposable parameters in the Read-Shockley equation, and, moreover, hardly any data were obtained in the range $0 < \theta < 8^\circ$ where the theory is expected to be most reliable.

Recently, there have been two studies of grain boundary energies in solids which are more complete and provide data an order of magnitude more precise than the earlier work. Gjostein (1958; also Gjostein and Rhines, 1958) has made detailed measurements of γ for (100) tilt and twist boundaries and showed that the Read-Shockley equation accurately agrees with experimental values of energy up to 5° to 6° misorientation, where the dislocations are spaced 10 atom planes apart. When the dislocations are closer together (i.e., for large θ values), the linear elasticity theory used by Read and Shockley can no longer account for the entire energy

¹ The common nomenclature in this field is to designate the angle of misorientation of the two crystals by θ and the orientation of the boundary relative to some principal axis as ϕ .

of the boundary. Wagner and Chalmers (1960) working with precisely aligned germanium tricrystals, have shown that (100) and (111) tilt boundaries fit the Read-Shockley equation extremely well from very low angles up to about 15° and 12° , respectively, and that γ then remains essentially constant for larger values of θ . Thus, the apparent and embarrassingly broad range of validity of the Read-Shockley equation suggested by the early work has now been satisfactorily resolved.

Wagner and Chalmers also found that the energy, γ , depends markedly on ϕ , the orientation of the boundary, for low-angle, but not for high-angle boundaries. This is the first time any such observation has been reported; according to Wagner and Chalmers, the dependence on ϕ may have become observable because of the precise alignment of their tricrystals, although a basic difference between germanium and metals in this respect has not been ruled out.

Important Problems—Since the dislocation model of a low-angle boundary is well established, the problems of major interest here are those relating to the detailed atomic configuration in the vicinity of the core of a dislocation. For instance, the effective self-diffusion coefficient in the neighborhood of a dislocation poses a problem which has neither been analyzed theoretically nor studied in detail experimentally, yet the mobility of atoms in this region is of fundamental importance in a variety of solid state reactions or processes. Another little understood phenomenon of broad interest, is the interaction between solute atoms and dislocations. Knowledge of such interactions is basic to any understanding of segregation or nucleation at, and diffusion along dislocations.

As far as problems which are more obviously related to grain boundaries are concerned, studies similar to those of Gjostein for low-angle twist boundaries, or for rotations about other principal axes, on a variety of different materials would certainly contribute to the understanding of dislocation structure and dislocation interactions in crystals. Another possibly fruitful project might be to study the variation of γ with ϕ for metals in as precise a manner as Wagner and Chalmers have done for germanium.

5.1.2. High-Angle Grain Boundaries

Low-angle boundaries appear to be quite simple but they occur relatively rarely in nature. The vast majority of boundaries found in annealed metals are classified as "high angle," that is, boundaries separating grains whose misorientation is large. The borderline

between "high" and "low" angle boundaries is not sharp, but, as we have seen, the calculation made by Read and Shockley becomes inapplicable in copper for $\theta > 6^\circ$. For tilt boundaries with $\theta \leq 19^\circ$ (dislocations 2-3 planes apart), the dislocation model is satisfactory but the calculation of the energy is extremely difficult. For still greater values of θ , the model loses its validity. No satisfactory *a priori* model exists for high-angle grain boundaries; experimental work in this field has consisted largely of studying various phenomena which show a variation with θ and ϕ and then inferring details about the structure of the boundary from such measurements.

(a) *Energy Measurements*—As mentioned above, γ has been measured for a variety of metals at large θ . In all cases, γ increases quite rapidly with θ at first and then approaches an essentially constant value for $\theta < 20^\circ$ (Gjostein, 1958). This lack of variation in γ at high angles suggests that the techniques employed may not be sensitive enough to distinguish between various boundaries at higher misorientations. One problem which has been studied with this technique is the presence of cusps or sharp relative minima in the γ vs. θ plots, corresponding to boundaries which are planes of symmetry and which contain a large number of lattice points belonging to both crystals. Twin boundaries are the best-known examples and the low energy of such boundaries is well documented. However, Read and Shockley (1950) have argued that a large number of energy cusps should exist. Gjostein and Wagner both looked carefully for cusps in their work on high-angle tilt boundaries, but failed to detect any aside from those associated with coherent twin boundaries. The reasons for this are probably similar to those for the blunting and disappearance of cusps on all but close-packed faces on free surfaces at moderate temperatures (Herring, 1953R; also sec. 2.4).

(b) *Diffusion in Grain Boundaries*—Evidence for preferential diffusion along grain boundaries has frequently been obtained (Amelinckx and Dekeyser, 1959R) but the interpretation of grain boundary diffusion data in terms of structural models is difficult. The methods currently available for analyzing grain boundary diffusion data do not yield absolute values of grain boundary diffusion coefficients. Instead, only a quantity δD_{gb} can be obtained from tracer diffusion experiments, D_{gb} being the grain boundary diffusion coefficient and δ an assumed "thickness" of the boundary. This quantity can only be crudely related to the actual structure of the boundary.

Some fairly systematic measurements have been carried out,

however, for simple tilt boundaries and for high-angle tilt boundaries (corresponding to rotation of the neighboring crystals about an axis lying in the plane of the boundary), particularly in metals such as silver and lead. The available data for silver may be summarized as follows:

(1) For $\theta < 15^\circ$, the grain boundary diffusion coefficient associated with atomic migration along the edge dislocation lines or "pipes" which make up the boundary increases linearly with θ (Turnbull and Hoffman, 1954). The largest part of this variation in grain boundary diffusion coefficient with angle may be attributable to the change in dislocation spacing with θ , since it appears that the diffusion coefficient along an individual dislocation pipe is practically independent of θ . Moreover, D_{gb} in a direction perpendicular to the dislocation pipes is roughly an order of magnitude lower than D_{gb} along the pipes (Hoffman, 1956).

(2) For $\theta > 15^\circ$, D_{gb} increases more rapidly with θ than a linear extrapolation of the data for $\theta < 15^\circ$ would indicate, and a maximum in D_{gb} is observed at $\theta \cong 45^\circ$. The observations seem to be consistent with the idea that high-angle boundaries can no longer be described in terms of simple dislocation arrays. The maximum in D_{gb} as a function of θ exhibits a smaller half-width than does the plot of γ vs. θ for boundaries of the same type (Hoffman, 1956; Couling and Smoluchowski, 1954).

(3) The grain boundary diffusion coefficient still appears to be anisotropic with direction even when $\theta = 45^\circ$ (Turnbull and Hoffman, 1945; McLean, 1957R).

As yet, diffusion studies of a systematic nature have not been carried out with twist-type boundaries, i.e., boundaries for which the axis of relative rotation is normal to the plane of the boundary. For relatively small rotations, boundaries of this type are believed to consist of crossed grids of screw dislocations (Read, 1953R; Amelinckx and Dekeyser, 1959R). It is known qualitatively that D_{gb} does vary with ϕ , for a given θ (Leymonie and Lacombe, 1957) but whether D_{gb} increases or decreases as the screw component of the boundary increases has not been established. Diffusion experiments on twist boundaries would be valuable, both in providing a detailed comparison with the data for tilt boundaries and possibly in helping to explain the very high mobilities observed for certain high-angle boundaries (sec. 5.3).

(c) *Structure of High-Angle Grain Boundaries*—Existing models are highly qualitative and although they allow the rationalization of the observed data on diffusion (Couling and Smoluchowski, 1954), it is difficult to make useful predictions from such models.

Observations of grain boundaries in bubble rafts give a lucid picture of how the separated dislocations of low-angle grain boundaries come closer together as θ increases, and how the dislocations finally lose their individual identity at $\theta \geq 20^\circ$ (Lomer and Nye, 1952). To the extent that metals can be described by nearest-neighbor interactions, the behavior of grain boundaries in metals should approximate that seen in the bubble raft. Thus, a model of sorts is not difficult to acquire. However, even assuming that the disordered region at the boundary is very narrow and using a simple interaction potential, calculations of the exact atomic positions at the boundary or of the interfacial tension encounter two serious difficulties. First, the problem is many-bodied, so that an exceedingly large number of simultaneous equations must be handled. Second, the temperature may have a pronounced effect on the structure of the grain boundary and this must somehow be taken into consideration. At the high temperatures where γ is measured, the entropy could decrease γ appreciably. The only possible way of treating the first problem is to make the calculations for some repeatable unit of finite size (Gjostein, 1958; Friedel, Cullity and Crussard, 1953). The marked effect of temperature on structure has been inferred by Fukushima and Ookawa (1957) using a dynamic bubble raft model. An actual calculation of the grain boundary entropy at this time would seem to be prohibitively difficult, however.

(d) *Important Problems* — Despite the difficulties involved, attempts at a more detailed calculation of the atomic structure of a high-angle grain boundary should be encouraged. If a suitable model could be developed, which would allow definite predictions to be made, this would provide a great stimulus to further experimental work in this field.

Diffusion data for twist-type boundaries are entirely lacking and would be useful both in interpreting structural models of high-angle boundaries and the mobilities of such boundaries.

Reliable data on the temperature dependence of γ for grain boundaries are nonexistent, but could be obtainable with currently available techniques. This would give an experimental value of the grain boundary entropy and complement the work on structure.

Such data as currently exist indicate that the energies of inter-phase boundaries in multicomponent systems are generally less than grain boundary energies in single phase systems. No explanation has yet been given for this rather unexpected result.

5.2. Grain Boundary Adsorption of Solute

Up to this point, only one-component systems have been considered. However, even in metals of high purity, the adsorption or segregation of solute atoms at grain boundaries appears to play an exceedingly important role. Impurity segregation is believed to have a pronounced effect on grain boundary mobility, the susceptibility to grain boundary fracture, the nucleation of second phase constituents at grain boundaries and many other related phenomena (McLean, 1957R; Amelinckx and Dekeyser, 1959R; Gifkins, 1959R).

5.2.1. Physical Models

Two physical effects which are usually considered to give rise to adsorption are strain energy interactions and electronic interactions (Cottrell, 1954). Qualitatively, they may be described as follows: if a solute atom is larger or smaller than the solvent atom, it will strain the lattice. This strain energy can be reduced if the solute atom moves to the grain boundary where it can find a hole whose diameter is commensurate with its own. If a solute atom has a different "valence" than the matrix atoms, it can be thought of as a charged ion. This ion will interact with the inhomogeneous electric field associated with the disordered structure of the grain boundary. Quantitative work, especially on the latter effect, must await a more detailed model of a high-angle grain boundary.

5.2.2. Techniques for Measuring Grain Boundary Adsorption

The experimental difficulties associated with quantitative measurements of grain boundary adsorption are immense. The procedure attempted most frequently in the past is to analyze directly for the excess solute associated with a given grain boundary. The difficulties encountered can be appreciated when it is considered that a single atomic layer of pure solute sets a rough upper limit on the amount of impurity that can be adsorbed, while the smallest element of thickness at the boundary which can be sampled with microchemical techniques currently available (see Part XII on "Techniques and Instrumentation," sec. 2.8) involves at least 10^4 atom layers.

Thomas and Chalmers (1955) have shown autoradiographically that polonium segregates to grain boundaries in lead but the quantitative interpretation of their results has since been questioned (Ward, 1957). More recently, Inman and Tipler (1958) have dis-

solved successive layers parallel to the grain boundary in a bicrystal of iron and have chemically analyzed for the segregation of phosphorus at the grain boundary. They report so much segregation that the analyses would correspond to the presence of a pure phosphorus film from 6 to 50 Å thick at the grain boundary. This conclusion is distasteful for several reasons (Cahn and Hilliard, 1959).

Nevertheless, recent studies of iron-sulfur alloys by Ainslie et al. (in press), in which the technique of thin-film transmission electron microscopy has been employed, have provided strong evidence for the segregation of sulfur near grain boundaries. It further appears that solute atoms diffusing in from grain boundaries can contribute to the formation of dense dislocation networks on either side of the boundary and thus act to pin or immobilize the boundary. While it would be difficult at present to adapt the technique of thin film microscopy to obtain quantitative measurements of grain boundary adsorption or segregation, there is no doubt that this technique will prove extremely valuable in the future, since it permits the detailed structural effects associated with impurity segregation at grain boundaries to be studied on a scale of resolution far exceeding that obtainable by any other method.

An alternative procedure, which avoids sampling of the grain boundary constituents (for direct analysis by chemical, autoradiographic, X-ray fluorescence, or other techniques) is to measure the variation of γ with solute concentration and then to use the Gibbs adsorption equation to establish the surface excess. In dilute solutions, the interpretation is reasonably straightforward. However, absolute values of the grain boundary energies or interfacial tensions are required and such data are difficult to obtain. The only serious attempt of this kind so far reported is that of Hilliard, Cohen and Averbach (1960).

5.2.3. Effects Due to Adsorption

There are a number of phenomena in which the adsorption or segregation of impurities at grain boundaries is thought to play a major role. Few if any of these effects are understood in detail at the present time. It is conceivable, however, that with additional effort some of the observed effects could be exploited and possibly developed into useful tools for studying the structure of grain boundaries or evaluating the extent of grain boundary segregation.

It is well established, for example, that as little as 10^{-4} or 10^{-5} weight percent of some solutes can markedly change the mobility

(i.e., the velocity per unit driving force) of certain grain boundaries in metals such as aluminum and lead (Aust and Rutter, 1959a, 1959b, 1960 and sec. 5.3 below). The pronounced effects which are observed even when only trace quantities of impurities are present would seem to require adsorption of the impurity atoms at grain boundaries.

It is also known that grain boundary relaxation phenomena, such as are observed by internal friction techniques, are extremely sensitive to the presence of trace impurities, which presumably are segregated at grain boundaries. However, the interpretation of the observed relation between peaks and activation energies is still subject to considerable uncertainty (see Part VIII on "Mechanical Behavior of Solids," sec. 4.3) and the orientation dependence of these relaxation effects has not yet been investigated.

The susceptibility to grain boundary fracture in many alloys at ordinary temperatures is frequently enhanced by slow cooling from elevated temperatures as well as by the addition of certain solutes. In dilute copper-base alloys, for example, where it is unlikely that second phase constituents could be present, adsorption at grain boundaries is believed to be responsible (McLean, 1957R).

In impure metals, melting occurs at the grain boundaries at temperatures somewhat below the melting point of the pure bulk solid. The effect is dependent upon the impurity concentration and may not be present in super-purity metals, although analogous effects apparently can occur at free surfaces even in the absence of impurities (Cahn and Hilliard, 1959b). As Weinberg (1958) has shown, working with carefully prepared bicrystals, grain boundary melting phenomena also exhibit a pronounced orientation dependence.

Another spectacular effect is that reported by Arkarov, Klotsman and Timofeev (1957), who found that small amounts of antimony have a pronounced influence on the extent of diffusion or penetration of silver into grain boundaries in copper. When only as little as 0.25% Sb is added to the silver, the shape of the plot of penetration depth vs. θ is changed and the depth of penetration of silver along some grain boundaries in copper is increased by as much as an order of magnitude.

5.2.4. Summary

Our present understanding of adsorption at grain boundaries and of various effects attributed to adsorption is quite poor, partly because a detailed structural model for high-angle grain boundaries is lacking and because quantitative data are nonexistent. The

most urgent need appears to be the development of an accurate technique for measuring the extent or amount of grain boundary adsorption. In our opinion, the technique which currently seems to offer most promise is that based on the use of the Gibbs adsorption equation, although more direct methods such as thin-film transmission electron microscopy may also prove of value in this connection. If a reliable procedure could be developed and applied even to a single system, many existing uncertainties could no doubt be resolved. In particular, it would be most desirable to obtain quantitative data on the extent of adsorption as a function of temperature, misorientation across the grain boundary, and relative valences and atomic sizes of solvent and solute atoms.

5.3. Grain Boundary Mobility and Migration

The factors influencing crystal boundary migration are treated in detail in Part VII, "Growth, Structure, and Morphology of Crystals," section 4.4. Here we will discuss only those experiments dealing with the mobility of grain boundaries which imply a variation of grain boundary structure with varying θ and ϕ . In theoretical studies of the atomic mechanism of grain growth, the quantity which is usually calculated is the mobility, m , or the velocity per unit driving force. Thus, to test the theory experimentally, a knowledge of both the velocity and the driving force is required. There are only one or two known techniques for determining absolute values of m (Mullins, 1956) and quantitative data are virtually nonexistent. However, data on the relative mobility of boundaries of different orientations can be obtained if an unknown but constant driving force is operative. A few studies of this type have been carried out.

The most outstanding work in this area is that of Aust and Rutter (1959a, 1959b, 1960), who have studied the mobility of grain boundaries in high-purity lead as a function of tin concentration and orientation. Some of their results may be summarized as follows:

- (1) Boundaries separating crystals with "special" orientation relationships move at velocities up to 100 times greater than boundaries separating crystals with "random" orientation differences. These special misorientations are those involving relative rotations about a common $\langle 111 \rangle$ or $\langle 100 \rangle$ axis.
- (2) The mobility of certain boundaries is decreased by a factor of 1000 as the tin concentration is increased from 4×10^{-4} to 6×10^{-3} weight per cent.
- (3) The ratio of the mobility of special boundaries to that of

random boundaries increases from approximately 1 at 10^{-5} weight per cent to 100 at 3×10^{-3} weight per cent Sn.

The first observation seems to apply to boundaries of all misorientations between the "special" lattice orientations. Kohara et al. (1958) have further shown that in 99.99% Al, the mobility of the boundaries separating two grains which are crystallographically related by a 40° rotation about a common $\langle 111 \rangle$ axis varies with the orientation, ϕ , of the boundaries, with "screw" type boundaries having a lower mobility than "edge" type.

The mobility, at least of certain boundaries, thus seems to depend upon both lattice misorientation and boundary orientation. It is also evident that impurities exert a pronounced effect and that the influence of impurities on grain boundary mobility must be established before we can arrive at a satisfactory description of the effect of grain boundary structure per se on boundary migration rates. At the present time, this whole field is in a state of flux but work of the type initiated by Aust and Rutter, now in progress, should add materially to our knowledge in this area.

5.4. References

5.4.1. Reviews and General Articles

- Amelinckx, S., and Dekeyser, W., 1959, in *Solid State Physics*, F. Seitz and D. Turnbull, editors (New York, Academic Press), pp. 327-491.
- Cottrell, A. H., 1954, in *Relation of Properties to Microstructure* (Cleveland, Ohio, American Society for Metals), p. 131.
- Gifkins, R. C., 1959, in *Fracture*, B. L. Averbach, et al., editors (Technology Press, MIT and John Wiley and Sons, New York), pp. 579-623.
- Herring, C., 1953, in *Structure and Properties of Solid Surfaces*, R. Gomer and C. S. Smith, editors (Chicago, University of Chicago Press), pp. 1-81.
- McLean, D., 1957, *Grain Boundaries in Metals* (Oxford, Oxford University Press).
- Read, W. T., 1953, *Dislocations in Crystals* (New York, McGraw-Hill).
- , and Shockley, W., 1952, in *Imperfections in Nearly Perfect Crystals*, W. Shockley, et al., editors (New York, John Wiley and Sons), pp. 352-376.

5.4.2. Original Articles

- Ainslie, N. G., Hoffman, R. E., and Seybolt, A. U., *Acta Met.*, in press.
- Arkarov, V. I., Klotsman, S. M., and Timofeev, A., 1957, *Phys. Metals, Metallography*, 5 (2), 152.
- Aust, K. T., and Rutter, J. W., 1959a, *Trans. Met. Soc. AIME*, 215, 119; 1959b, *ibid.*, 215, 820; 1960, *ibid.*, 218, 682.
- Cahn, J., and Hilliard, J. E., 1959a, *Acta Met.*, 7, 219.
- , and ———, 1959b, *J. Chem. Phys.*, 31, 688.
- Couling, L. R. S., and Smoluchowski, R., 1954, *J. Appl. Phys.*, 25, 1538.
- Friedel, J., Cullity, B., and Crussard, C., 1953, *Acta Met.*, 1, 80.

- Fukushima, E., and Ookawa, A., 1957, *J. Phys. Soc. (Japan)*, **12**, 139.
- Gjostein, N. A., 1958, Thesis, Carnegie Institute of Technology, Pittsburgh, Pa.
- Gjostein, N. A., and Rhines, F. N., 1959, *Acta Met.*, **7**, 319.
- Haynes, C. W., and Smoluchowski, R., 1955, *Acta Met.*, **3**, 130.
- Hilliard, J. E., Cohen, M., and Averbach, B. L., 1960, *Acta Met.*, **8**, 26.
- Hoffman, R. E., 1956, *Acta Met.*, **4**, 97.
- Inman, M., and Tipler, H., 1958, *Acta Met.*, **6**, 73.
- Kohara, S., Parthasarathi, M. N., and Beck, P. A., 1958, *Trans. AIME*, **212**, 876.
- Leymonie, C., and Lacombe, P., 1957, *Rev. Met.*, **54**, 653.
- Li, C. H., Edwards, E. H., Washburn, J., and Parker, E. R., 1953, *Acta Met.*, **1**, 223.
- Li, J. C. M., 1960, *Acta Met.*, **8**, 296.
- Lomer, W., and Nye, J. F., 1952, *Proc. Roy. Soc. (London)*, **A212**, 576.
- Mullins, W. W., 1956, *Acta Met.*, **4**, 421.
- Read, W. T., and Shockley, W., 1950, *Phys. Rev.*, **78**, 275.
- Thomas, W. R., and Chalmers, B., 1955, *Acta Met.*, **3**, 17.
- Turnbull, D., and Hoffman, R. E., 1954, *Acta Met.*, **2**, 419.
- van der Merwe, J. H., 1950, *Proc. Phys. Soc. (London)*, **A63**, 616.
- Vogel, F. L., Pfann, W. G., Corey, H., and Thomas, E., 1953, *Phys. Rev.*, **90**, 489.
- Wagner, R. S., 1959, Thesis, Harvard University, Cambridge, Mass.
- , and Chalmers, B., 1960, *J. Appl. Phys.*, **31**, 581.
- Ward, R. G., 1957, *Acta Met.*, **5**, 681.
- Weinberg, F., 1958, *Acta Met.*, **6**, 535.

6. FRICTION, ADHESION AND LUBRICATION

6.1. Friction

6.1.1. Simple Adhesion Theory of Metallic Friction

Coulomb first suggested that friction might be due to the interlocking of surface asperities and this idea still receives some support. There is, however, a strong body of experimental evidence (Bowden and Tabor, 1954R) which suggests that the friction of solids is primarily due to the strong adhesion that occurs at the regions of real contact.

This mechanism first evolved from studies on metal surfaces. Such surfaces, however carefully prepared, are rough on the atomic scale, and when placed on one another touch only at the tips of their asperities. The elastic limit is apparently exceeded for even the smallest loads and plastic flow occurs. The plastic yield pressure of a metal over a small localized region depends but little on the shape or geometry of the deformed region. To a first approximation, the plastic yield pressure may be treated as a material constant, p_0 , somewhat analogous to the indentation hardness of the metal. Consequently, the asperities crush down plastically until

they are large enough to support the applied load W . Hence, the true area of contact is $A = W/p_o$, with A being proportional to the load but independent of the size of the surfaces. At these regions of contact, strong adhesion occurs even at room temperature (cold welding). This applies both to clean surfaces and to surfaces covered with contaminants which can be penetrated or disrupted by the sliding process. The junctions so formed are probably quite small; thus, some electrical resistance measurements by Holm (1946R) suggest diameters of the order of 10^{-4} cm.

On this view, the friction may be described quantitatively as the force required to shear the junctions formed at the interface. In addition, if one surface is harder than the other, the asperities of the harder surface will plough their way through the softer one. This contribution is generally small, so that we may write $F = As_o$, where s_o is the specific shear strength of the junction. The coefficient of friction then becomes $\mu = F/W = s_o/p_o$.

This treatment has been reasonably successful in explaining Amontons' or Coulomb's two laws of friction (i.e., F is proportional to W , and F is independent of the size of the surfaces), as well as the small effect of hardness and temperature on friction, the lubricating effect on soft metallic films, and many other observations on frictional phenomena.

6.1.2. Friction of Non-Metals

The simple adhesion theory of friction has also proved surprisingly useful in explaining the friction of non-metals. For most such materials a similar adhesion mechanism applies, the main difference being in the way the true area of contact depends upon the load and upon the geometry of the surfaces. For non-metals, the most marked divergence occurs if the surfaces deform elastically. If we consider a single spherical asperity, the area of contact is proportional to $W^{2/3}$. If, however, the effect of increasing the load is to bring new asperities into contact, the index will be greater than $2/3$ so that A will approach linear proportionality with W . If the shear strength of the interface remains constant, the friction will vary with load in a similar way (Tabor, 1957; Archard, 1957; Schallamach, 1958).

A more elaborate treatment, designed to cover the behavior of solids in general, has been given by Rubenstein (1956) and this has been used to explain the observed frictional behavior on the basis of the adhesion mechanism. It should, however, be borne in mind that such refinements in the analytical treatment impose a heavy burden on the fundamental mechanism.

Another problem may arise with brittle solids. With many such materials, the compressive stresses around each individual asperity appear to inhibit brittle fracture, and the deformation and shearing processes are often dominated by the plastic rather than the brittle properties of the material; as a result the frictional behavior resembles that of metals (King and Tabor, 1954). However, with excessively brittle solids, fragmentation probably plays a major part in the sliding process. Apart from these problems concerning the way in which the surface asperities deform, there is a third issue which may become of considerable importance. With some nonmetals — especially viscoelastic materials — the ploughing or deformation term may make a relatively large contribution to the total frictional force. Even if the deformations are elastic, appreciable energy losses may occur as a result of the hysteretic properties of the material. This can be particularly marked if the surface adhesion is intrinsically low or is reduced deliberately by the use of lubricant films (Greenwood and Tabor, 1958).

6.1.3. More Critical Theory of Metallic Friction

One of the major inadequacies of the simple adhesion theory is that it treats the yield pressure, p_0 , and the specific shear strength, s_0 , of the interface as two independent quantities. This cannot be so. The yielding at the interface occurs as a result of the combined normal and tangential stresses. When the surfaces are first placed in contact the asperities deform plastically under the influence of the applied load. The smallest tangential stress should thus be able to produce further (irreversible) plastic flow with a corresponding increase in the area of contact. This has been observed; for example, as the tangential force is increased from zero to a value less than that at which gross sliding occurs, there is a minute tangential displacement accompanied by steady junction-growth (Parker and Hatch, 1950; McFarlane and Tabor, 1950) and the behavior may be represented approximately by a relation of the form $p^2 + \alpha s^2 = p_0^2$, where p and s are the normal and tangential stresses at any instant in the junction (McFarlane and Tabor, 1950; Courtney-Pratt and Eisner, 1957). With thoroughly outgassed metals, junction-growth generally proceeds until the whole of the geometric area is in contact, coefficients of friction of the order of 50 to 100 being observed (Bowden and Young, 1951; Gwathmey, et al., 1952). It would now seem that, in the presence of contaminant films, the initial stages of junction-growth are essentially the same as for clean surfaces. However, as soon

as the applied shear stress exceeds the critical shear strength of the interface, junction-growth comes to an end and macroscopic sliding occurs within the interfacial layer itself. Thus, if the interface is only slightly weaker than the surrounding bulk metal, the combined stress equation shows that sliding should occur when μ is of the order of unity (Tabor, 1959). This is in good agreement with the experimental observation (Bowden and Young, 1951) that if a minute amount of oxygen or water vapor is admitted to outgassed surfaces, gross sliding occurs at $\mu \cong 1$.

Another point of interest in any more detailed study of friction is the shape of the asperities and the way in which they interact. Interlocking of asperities, as suggested by Coulomb and later workers (Ming-Feng, 1954), does not often contribute significantly to the frictional resistance. However, the way in which one asperity encounters another may critically affect the geometry of the junction which is formed and this may have a marked influence on its subsequent behavior; for example, the junction may shear plastically, or tear before shearing or even shave the top of one of the surfaces as sliding takes place (Green, 1955; Greenwood and Tabor, 1955, 1957). This must affect not only the friction but also the wear of the surfaces.

6.1.4. Friction and Chemical Decomposition

The friction and impact of solids may initiate chemical decomposition. An important factor in these tribochemical effects is the high temperature flashes developed at the points of contact (Bowden and Yoffe, 1952R).

6.1.5. Rolling Friction

When one body rolls freely over another, a minute amount of interfacial slip occurs. This process has now been analyzed both theoretically and experimentally and a good deal more is known about the mechanism than was true as recently as ten years ago (Johnson, 1958). However, in "free" rolling (where the rolling element does not transfer any tangential traction) interfacial slip generally contributes little to the observed rolling friction. The greater part of the friction arises from hysteresis losses in the bulk of the material (Tabor, 1955; Eldredge and Tabor, 1955). For this reason, rolling friction is not greatly affected by the presence of lubricant films. Hysteresis losses are very small for hard steels such as are used in ball and roller bearings; they can, however, be substantial in polymeric solids and rubbers (Greenwood and Tabor, 1958; Bueche and Flom, 1959). Hysteresis losses not only play a

large part in rolling friction, but they also have some influence on the fatigue life of the materials.

6.1.6. Problems in Friction

Range of Action of Surface Forces—This problem is fundamental to both friction and adhesion. Theoretical studies show that the force between molecules for van der Waals interaction falls off as x^{-7} for small separations, x , and as x^{-3} for relatively large separations (Casimir and Polder, 1948; Lifschitz, 1954). For extended surfaces, the interaction force of all the molecules on one surface with all those on the other may be calculated. For a hemispherical surface resting on a flat, it turns out that the total force (for larger separations) falls off as H^{-4} , where H is the distance of closest approach (Lifschitz, 1955). Experimental studies have confirmed the existence of surface forces which fall off in approximately this way (Abrokosova and Derjaguin, 1957), although there are some conflicting data (Overbeek and Sparnaay, 1954).

Most of the work has been carried out with highly polished surfaces; consequently, the distance of minimum separation, H , is of the order of 1,000Å and the forces observed are of the order of 10^{-6} gm. weight. Experimental difficulties are formidable in this range and are complicated by the enormous effect of stray electrostatic charges or the presence of minute dust particles (Sparnaay, 1957). It may be noted that a net force of 10 gm. is to be expected only when the separation is less than about 100Å. In this more interesting range no direct experiments have yet been made though they would be of major interest. The best indirect estimate of surface forces very close to the solid surface has been based on the infra-red spectra of molecules adsorbed on silica glass (Sheppard and Yates, 1956). Certain characteristic features appear in the spectra only when the molecule is subjected to a large electrostatic field. This field has been estimated to be of the order of 7×10^6 volts/cm.

Surface States as Affected by Mechanical Deformation—When metal surfaces are disturbed mechanically it is found that they often emit charged particles (Kramer effect). In the presence of moisture this may lead to the formation of hydrogen peroxide (Russell effect). Earlier work provided little understanding of the mechanism involved but it is now apparent that the charged particles are electrons; mechanical deformation appears to produce centers where the electrons are loosely attached and can be readily liberated by exposure to light or other sources of energy (Grun-

berg and Wright, 1955). A helpful general survey of this field has recently been published (Grunberg, 1958R). It is clear that the surface energy states produced by mechanical working must play an important part in surface oxidation, in catalysis [see in particular the recent work of Hermance and Egan (1958)], and in the reactivity of the surface with lubricants, but much remains to be done before such effects can be clarified.

Adhesion and Strength Properties—A more detailed understanding is required of the way in which minute interfacial junctions are formed and sheared, in terms of modern dislocation concepts (Cottrell, 1953R; Read, 1953R).

Junction formation: It appears that one important parameter is the time available for the surfaces to come into intimate contact and for adhesion to occur. Why this should be so is not understood.

Shearing of junctions: There seem to be several important variables; (a) the rate of shearing, since strength properties, at least of bulk materials, depend upon the strain rate; (b) the size of the junctions, since below certain dimensions the size has a profound effect on strength and on the mode of yielding; (c) surface films which may envelop the junction, since such films can also greatly affect mechanical properties; (d) pressure on the junction. Some workers (Lodge and Howell, 1954; Kraghelsky and Sabelnikov, 1957; Rubenstein, 1958) suggest that the specific shear strength is dependent on the pressure over the region of contact, citing the work of Bridgman (1952R) in support, but the effect requires further study.

Oxide Films—Under most conditions, the frictional behavior is dominated by the oxide "film" on the surface. In the past this has been regarded as a uniform film separating the surfaces (Whitehead, 1950; Cocks, 1952). Recent work (Pfefferkorn, 1953; Gulbransen and Copan, 1959) shows that the oxide is much more complex and often grows in the form of fine crystallites, the detailed morphology depending on the orientation of the substrate and the oxidizing conditions. It would be very valuable to interpret the role of surface oxides in friction with this newer information in mind.

Measurement of Area of True Contact—Electrical methods have been used for metals (Holm, 1946R; Bowden and Tabor 1954R), optical for nonmetals (Dyson and Hirst, 1954; Bailey and Courtney-Pratt, 1955). Existing techniques are not really satisfactory and often blur over the crucial evidence as to the type of surface deformation occurring. An unequivocal method for measuring the true area of contact would be of great value.

Elastic Area of Contact—Some workers (Blok, 1952) suggest that, with very well matched hard steel surfaces or even with well run-in brass surfaces, the area of contact may be determined by elastic rather than plastic yielding. This can presumably be true during sliding only if the interfacial adhesion is very weak, i.e., only in the presence of superficial oxide or lubricant films. The problem merits further study.

Junction Growth—A better plasticity model is needed to describe junction growth. Other factors which may influence the termination of junction growth should be considered, e.g., ductility and work hardening (Rubenstein, 1958; Tabor, 1959). Friction experiments carried out using freshly cut surfaces in high vacua would be desirable for comparison with data on outgassed metals, which are inevitably annealed. In addition, more detailed knowledge is required of the way in which encountering asperities interact.

Intermittent or "Stick-Slip" Motion—This arises when the friction diminishes with increasing speed. A great deal of effort has been expended on the analysis of stick-slip motion as a problem in mechanics (Derjaguin, Push and Tolstoi, 1957), but the reason for the decrease in friction with time of contact (Rabinowicz, 1958) has still not been clearly established.

Temperature Flashes—Temperature flashes between sliding surfaces have been studied theoretically (Blok, 1937; Jaeger, 1942) and experimentally (Bowden, Stone and Tudor, 1947; Parker and Marshall, 1948; Bowden and Thomas, 1954). The theoretical analysis at very high speeds (where the heat penetration during the transit time of any individual asperity is very small) needs further elaboration, although a start in this direction has been made (Bowden and Freitag, 1958).

Effect of Sliding on Surface Structure—Sliding produces deformation, crystal rotation (Scott and Wilman, 1958), surface flow, etc. The structural changes which occur at the interface have yet to be analyzed in detail.

6.2. Adhesion

6.2.1. Thermodynamics of Adhesion

The phenomenon of adhesion is frequently interpreted in terms of surface energies. Although this is a valid approach for liquid-liquid and possibly some liquid-solid interfaces, where any work done may be equated to changes in free surface energy, it is unlikely that this same approach can be applied to solids. In pulling solids apart, the surface energy changes are often trivial compared

with the work expended in deforming the interface before failure occurs (Tabor, 1951R).

6.2.2. Adhesion of Metals

If two perfectly clean pieces of metal are placed in contact, the atoms at the interface are not conscious of their original affiliations and adhesion, similar in nature to the cohesion between the atoms in bulk, takes place. If the region of contact is altered by the application of comprehensive loads, the area of contact will be roughly proportional to the applied load. There is some evidence that the observed adhesion may be smaller than expected if the mutual orientation of the contacting elements is unfavorable (King and Tabor, 1954; Semenoff, 1958).

However, a far more important factor in causing low adhesion is the action of released elastic stresses, which may strip the junctions apart as the joining load is removed. This may be overcome by heating the specimens for a short period at a temperature of about $0.4 T_m$, where T_m is the absolute melting temperature of the material. The heat treatment anneals the junctions, thereby increasing their ductility and enabling them to survive the stripping action of the released elastic stresses (Bowden and Rowe, 1956). In addition, the junctions probably contain appreciable regions of lattice misfit which are potential sources of weakness and, at these temperatures, a sufficient amount of diffusion occurs to "heal" these imperfections.

Studies of the effect of time and temperature of heating suggest, however, that the primary effect of heating for short periods at $T \cong 0.4 T_m$ is to anneal or heat the junctions and not to increase their area of contact by sintering. Thus, with low-melting ductile metals, for which room temperature is greater than $0.4 T_m$, strong cold-welding can occur if the surfaces are reasonably clean (McFarlane and Tabor, 1950). With higher melting metals, strong welding can occur even below $0.4 T_m$ if the geometry of the surfaces is such as to minimize the effect of released elastic stresses (Mason, 1958). In general, however, a temperature around $0.4 T_m$ represents the minimum for satisfactory adhesion. At still higher temperatures, diffusion and surface tension effects are predominant (Herring, 1953R; Mackenzie and Shuttleworth, 1949); the area of contact increases with time and large scale sintering occurs.

6.2.3. Adhesion of Contaminated Metals

When metals are covered with oxide "films" or lubricant layers the adhesion is greatly reduced. These films may be penetrated by

subjecting the surfaces to relatively gross deformation (Cocks, 1954; Wilson, 1956). Under these conditions, provided the junctions survive the release of elastic stresses, strong adhesion will occur. Indeed, Anderson (1959) goes so far as to suggest that adhesion is observed only if shearing takes place. The simple view is to regard the breakdown of oxide films as depending mainly on their ductility relative to the underlying metal (Wilson, 1956). Hard oxides on soft metals are generally penetrated rather easily (tin oxide on tin). Ductile oxides may flow and be able to resist penetration (e.g., copper oxide on copper). Oxides on hard substrates are difficult to rupture unless the stresses are large enough to produce appreciable deformation of the substrate (oxides on hard steels). This simple and highly qualitative approach demands a critical reappraisal, particularly in view of our new knowledge of the complex morphology of oxides formed on metals (Gulbransen and Copan, 1959). In contrast to oxide films, lubricant films are often sufficiently mobile to spread over the surface; they may thus prevent adhesion even when the surfaces have been subjected to heavy deformations.

6.2.4. Adhesion of Non-Metals

The adhesion of non-metals follows the same general pattern as metals. Here, however, different types of bonding are available at the interface. For a solid in contact with itself the bonding will resemble that of the bulk. For different materials, bonding will depend on the interaction between say polar and nonpolar forces. With oxides (e.g., ZnO on NiO) it is probable that electron transfer at the interface plays an important part in the adhesion.

A similar mechanism has been suggested for the adhesion between a polymeric solid and a metal surface. Skinner, Savage and Rutzler (1953) have found that, when the polymer is stripped off the metal, an appreciable separation of electrostatic charge occurs if the break takes place truly at the interface. In their measurements, these charges could account for only a small fraction of the adhesive strength; under other conditions, however, e.g., in the cleavage of mica (Gaines and Tabor, 1956) charge separation may be of much greater significance. Indeed, Derjaguin and his associates (Derjaguin, 1955) maintain that electrostatic forces provide the greatest or most important contribution to adhesion. The subject merits further study.

With very inert materials such as Teflon, the adhesion is, by contrast, extremely weak.

6.2.5. Action of Adhesives

There are two main requisites for a successful adhesive. It must wet the surface, and when it solidifies it must not produce residual stresses due to contraction, expansion, etc. (de Bruyne and Houwink, 1951R). Apart from these two considerations, the chemical composition appears to be of secondary importance. While wetting clearly depends on interfacial tensions, the importance of wetting extends beyond purely thermodynamic considerations, since a non-wetting adhesive will tend to leave air pockets and unfilled crevices at the interface (Wake, 1950R) and these may give rise to local stress concentrations.

The behavior of "tacky" adhesives is different from that of adhesives which solidify. With "tacky" adhesives the adhesion is largely due to viscous forces involved in the act of separating the surfaces.

6.2.6. Problems in Adhesion

The basic problems in adhesion are those involving the energetics, structure and kinetic behavior of interphase interfaces. These problems are basically similar in nature to those encountered with grain boundaries, as discussed in section 5. There are few if any satisfactory methods for measuring adhesion quantitatively. Most methods tend to peel the surfaces apart or to introduce stress concentrations. The development of a suitable measuring technique offers a challenge to workers in the field.

An understanding of the adhesion of metals seems to hinge largely on such problems as the range of action of surface forces, the relation between adhesion and strength properties, and the influence of oxide films (sec. 6.1.6), all of which require elucidation. With respect to the adhesion of nonmetals, systematic studies of the relation between chemical structure and adhesion are still lacking. More work is needed on the role of electrical charge separation in adhesion. There is also a need for a thorough attack on the ancient problem of frictional electricity. A recent survey (Loeb, 1958R) emphasizes the complexity of the phenomena and the unsatisfactory state of this field, both theoretically and experimentally.

6.3. Lubrication

When a journal rotates in a lubricated bearing it drags a convergent wedge of oil along with it. The convergence generates a hydrodynamic pressure in the oil film and, if this is sufficiently large, the journal will "float" on the oil wedge. This corresponds

to fluid or hydrodynamic lubrication; friction arises only from the viscosity of the oil and there is, ideally, no contact between, and therefore no wear of, the sliding solids (Barwell, 1956R).

Hydrodynamic lubrication may also be obtained between parallel surfaces because of viscous heating of the oil. This causes thermal expansion and may, if the geometric arrangement is suitable, create an effective wedge action ("thermal wedge") between the sliding surfaces.

In practice, if speeds are too low or loads too high, the film necessary for hydrodynamic lubrication may become thinner than the dimensions of surface asperities. The solids will rub on one another and will be separated at most by a few molecular layers of lubricant. This is known as boundary lubrication (Hardy, 1936R). The friction does not depend on the bulk viscosity of the lubricant and slight but definite wear can occur. Boundary lubrication is best studied at high loads and low sliding speeds. It is then found that Amontons' laws of friction again apply. If pure hydrocarbons and their derivatives are studied, it is found that paraffins, alcohols and fatty acids lubricate most effectively when solid. A marked deterioration occurs at temperatures near the bulk melting point. Fatty acids, however, often provide effective lubrication at temperatures well above their bulk melting point. At one stage in the theory this was attributed to increased dipole-surface interaction. It is now believed that, with reactive metals, the acids form soaps which are effective up to their own higher melting temperatures. On nonreactive metals, fatty acids are scarcely more effective than paraffins (Bowden and Tabor, 1954R).

When unlubricated and boundary-lubricated surfaces are compared, it is found that, for solid boundary films, the friction is reduced by a factor of ~ 20 (from say $\mu = 1$ to $\mu = 0.05$) while the pick-up or transfer is reduced by a factor of 10^5 or more. At temperatures near the melting point, the friction increases by a factor of ~ 10 and the transfer by a factor of 10^2 to 10^3 . The lubricant still appears to be moderately effective. At somewhat higher temperatures, the lubricant is completely desorbed and the friction and surface damage are similar to those observed in the absence of lubricant. These changes correspond to changes in state of the lubricant film with temperature, so that they are reversible (Rabinowicz and Tabor, 1951).

Similar effects have been observed in electron diffraction studies of absorbed monolayers (Chapman and Tabor, 1957). Strong orientation of long-chain polar molecules disappears at a temperature very close to the bulk melting point of the monolayer mate-

rial. On cooling, orientation slowly returns. If heated to appreciably higher temperatures in a vacuum, the film is desorbed (Chapman and Tabor, 1957). Thus, an effective boundary lubricant is one which forms a coherent film separating the solid surfaces (to give a minimum of wear) and possesses a low shear strength (to give a low friction). It should remain solid up to high temperatures and be sufficiently polar to remain adsorbed at even higher temperatures.

6.3.1. "Extreme Pressure" Lubrication and "Solid-Film" Lubrication

Most organic materials suitable as boundary lubricants are totally desorbed at temperatures above about 250°C. Thus, under severe conditions of load and speed, local temperature flashes may render the lubricant completely ineffective. Under these conditions additives must be used which provide protective films up to higher temperatures. These are known as extreme pressure (or extreme-temperature) additives and usually consist of molecules containing active (labile) chlorine or sulphur which can react with the surfaces to form protective metal sulphides or metal chlorides.

Additives of this type also have their limitations, and, in some applications involving extremely high temperatures, solid films of such materials as graphite or molybdenum disulphide are used instead. These substances have a lamellar structure and exhibit pronounced basal cleavage. The cleavage planes correspond to low energy surfaces, the other crystal faces to high energy surfaces. There is usually sufficient adsorbed oxygen or water vapor on the other crystal faces to render them low energy surfaces too, so that the friction is generally low. For example, with graphite the adsorbed water may be driven off at high temperatures and the friction and wear may increase considerably (Savage, 1948). On the other hand, boron nitride appears to have poor lubricating properties until it has been slightly oxidized; apparently, the surface layers of boric oxide act at high temperatures as effective lubricant films between the boron nitride crystals (Deacon and Goodman, 1958). Because lamellar solids often have low energy cleavage surfaces it is very difficult to attach them to the surfaces of the sliding bodies. This is easier on softer bodies as a result of increased mechanical embedment (Deacon and Goodman, 1958).

6.3.2. Mixed Lubrication

Between the fluid and boundary regimes there is a very wide region known as the region of mixed lubrication. Many engines

and machine parts operate successfully under intermediate conditions of this kind where it is clear that classical hydrodynamic lubrication simply cannot apply. It appears that this is largely due to the vast increase in viscosity with pressure of most organic lubricants (Blok, 1951; Crook, 1958). If the lubricant is trapped between the moving surfaces, it behaves almost like a solid and so prevents the surfaces from coming into contact. The separating film is, however, extremely thin in engineering terms so that very good alignment and surface finish are necessary to exploit this effect (Crook, 1958).

6.3.3. Problems in Lubrication

An important general problem in the field of both hydrodynamic and boundary lubrication is the effect of surface active materials on the static strength and fatigue properties of solids (Rehbinder effect; Rehbinder and Lichtman, 1957; Andrade and Randall, 1952). The existing evidence is often highly conflicting and systematic studies under carefully controlled conditions are needed to resolve these inconsistencies.

Hydrodynamic Lubrication—There are still many unsolved theoretical problems involving boundary conditions, cavitation of lubricant, side leakage, and viscous heating. In addition, there are three major variables that are being intensively studied in many laboratories—the effect of pressure, temperature and rate of shear on the viscosity of the lubricant.

Boundary Lubrication—More information is needed about the orientation, structure, and two-dimensional “melting” of monolayers adsorbed on solid surfaces. Again, the physical properties of monolayers and the effect of pressure on the shear strength of monolayers require direct investigation (Bailey and Courtney-Pratt, 1955). In addition, there is need for a more systematic correlation of structure with lubricating properties.

Extreme Pressure and Dry Film Lubrication—The action of chlorine—and sulfur-containing additives is fairly well understood, but that of phosphate coatings is still partly mysterious. The mode of attachment of lamellar solids to sliding bodies and the way in which these anisotropic materials shear during sliding are both problems which require further attention.

Mixed Lubrication—The effect of pressure and temperature on viscosity are of major importance here (Crook, 1958). A study of effective lubrication as a function of pressure-viscosity characteristics is needed (Christopherson, 1957). It may also turn out that, at very high speeds, time effects may be of increasing importance

in determining the mechanical properties of the lubricant. This constitutes one part of the field of "elasto-hydrodynamic" lubrication.

6.4. References

6.4.1. Reviews and General Articles

- Barwell, F. T., 1956, *Lubrication of Bearings* (London, Butterworth's Scientific Publications).
- Bowden, F. P., and Tabor, D., 1954, *Friction and Lubrication of Solids* (Oxford, Clarendon Press).
- , and Yoffe, A. D., 1952, *Initiation and Growth of Explosions in Liquids and Solids* (Cambridge, Cambridge University Press).
- Bridgman, P. W., 1952, *Studies in Large Plastic Flow and Fracture* (New York, McGraw-Hill).
- Cottrell, A. H., 1953, *Dislocations and Plastic Flow in Crystals* (Oxford, Oxford University Press).
- deBruyne, N. A., and Houwink, R., 1951, *Adhesion and Adhesives* (New York, Van Nostrand).
- Grunberg, L., 1958, *Brit. J. Appl. Phys.* 9, 85.
- Hardy, Sir W. B., 1936, *Collected Works* (Cambridge, Cambridge University Press).
- Herring, C., 1953, in *Structure and Properties of Solid Surfaces*, R. Gomer and C. S. Smith, editors (Chicago, University of Chicago Press), pp. 5X72.
- Holm, R., 1946, *Electric Contacts* (Uppsala, Almqvist and Wiksells).
- Loeb, L. B., 1958, *Handbuch der Physik* (Berlin, Springer-Verlag).
- Read, W. T., 1953, *Dislocations in Crystals* (New York, McGraw-Hill).
- Tabor, D., 1951, *Ann. Rep. Progr. Appl. Chem.*, 36, 621.
- Wake, W. C., 1960, in *Adhesion*, D. D. Eley, editor (Oxford, Oxford University Press), Ch. VIII.

6.4.2. Original Articles

- Abrokosova, I. I., and Derjaguin, B. V., 1957, *Second International Congress of Surface Activity* (London, Butterworth's Scientific Publications), Part III, p. 398.
- Anderson, O. L., 1959, *J. Appl. Phys.*, 30, 593.
- Andrade, E. N. da C., and Randall, R. F. Y., 1952, *Proc. Phys. Soc. (London)*, B65, 445.
- Archard, J. F., 1957, *Proc. Roy. Soc. (London)*, A243, 190.
- Bailey, A. I., and Courtney-Pratt, J. S., 1955, *Proc. Roy. Soc. (London)*, A227, 500.
- Blok, H., 1937, "General Discussion on Lubrication," *Inst. Mech. Engrs.*, 2, 222.
- , 1951, *Proc. 3rd World Petrol. Congr.*, 7, 304.
- , 1952, *Proc. Roy. Soc. (London)*, A212, 480.
- Bowden, F. P., and Freitag, E. H., 1958, *Proc. Roy. Soc. (London)*, A243, 350.
- , and Rowe, G. W., 1956, *Proc. Roy. Soc. (London)*, A233, 429.
- , Stone, M. A., and Tudor, G. K., 1947, *Proc. Roy. Soc. (London)*, A188, 329.
- , and Thomas, P. H., 1954, *Proc. Roy. Soc. (London)*, A223, 29.
- , and Young, J. E., 1951, *Proc. Roy. Soc. (London)*, A208, 311.

- Bueche, A. M., and Flom, D. G., 1959, *Wear*, **2**, 168.
- Casimir, H. B. C., and Polder, D., 1948, *Phys. Rev.*, **73**, 360.
- Chapman, J. A., and Tabor, D., 1957, *Proc. Roy. Soc. (London)*, **A242**, 96.
- Christopherson, D. G., 1957, *Proc. of Conference on Lubrication and Wear* (London, Institution of Mechanical Engineers), p. 9.
- Cocks, M., 1952, *Nature*, **170**, 208.
- , 1954, *Proc. Phys. Soc. (London)*, **B67**, 238.
- Courtney-Pratt, J. S., and Eisner, E., 1957, *Proc. Roy. Soc. (London)*, **A238**, 529.
- Crook, A. W., 1958, *Phil. Trans. Roy. Soc.*, **A250**, 387.
- Deacon, R. F., and Goodman, J. F., 1958, *Proc. Roy. Soc. (London)*, **A243**, 464.
- Derjaguin, B. V., 1955, *Research*, **8**, 70.
- , Push, V. E., and Tolstoi, D. M., 1957, *Proc. of Conference on Lubrication and Wear* (London, Institution of Mechanical Engineers), p. 257.
- Dyson, J., and Hirst, W., 1954, *Proc. Phys. Soc. (London)*, **B67**, 309.
- Eldredge, K. R., and Tabor, D., 1955, *Proc. Roy. Soc. (London)*, **A229**, 181.
- Gaines, G. L., and Tabor, D., 1956, *Nature*, **178**, 1304.
- Green, A. P., 1955, *Proc. Roy. Soc. (London)*, **A228**, 191.
- Greenwood, J. A., and Tabor, D., 1955, *Proc. Phys. Soc. (London)*, **B68**, 609; 1958, *ibid.*, **B71**, 989.
- , and ———, 1957, 1958, *Proc. of Conference on Lubrication and Wear* (London, Institution of Mechanical Engineers), p. 314.
- Grunberg, L., and Wright, K. H. R., 1955, *Proc. Roy. Soc. (London)*, **A232**, 403.
- Gulbransen, E. A., and Copan, T. P., 1959, *Disc. Faraday Soc.*, **28**, 229.
- Gwathmey, A. T., Leidheiser, H., and Smith, G. P., 1952, *Proc. Roy. Soc. (London)*, **A212**, 464.
- Hermance, H. W., and Egan, T. F., 1958, *Bell System Tech. J.*, **37**, 739.
- Jaeger, J. C., 1942, *Journal and Proc. Roy. Soc. N.S.W.*, **76**, 203.
- Johnson, K. L., 1958, *J. Appl. Mech.*, **25**, 332.
- King, R. F., and Tabor, D., 1954, *Proc. Roy. Soc. (London)*, **A223**, 225.
- Kraghelsky, I. V., and Sabelnikov, V. P., 1957, *Proc. of Conference on Lubrication and Wear* (London, Institution of Mechanical Engineers), p. 246.
- Lifschitz, E. M., 1954, *Doklady Akad. Nauk. S.S.S.R.*, **97**, 643; 1955, *ibid.*, **100**, 879.
- , 1955, *Zhur. Eskptl. i. teoret. Fiz.*, **29**, 94.
- Lodge, A. S., and Howell, H. G., 1954, *Proc. Phys. Soc. (London)*, **B62**, 89.
- Mackenzie, J. K., and Shuttleworth, R., 1949, *Proc. Phys. Soc. (London)*, **B62**, 833.
- Mason, W. P., 1958, *Symposium* (Chicago, American Society of Lubrication Engineers), to be published.
- McFarlane, J. S., and Tabor, D., 1950, *Proc. Roy. Soc. (London)*, **A202**, 224.
- Ming-Feng, I., 1954, *Lubrication Eng.*, **10**, 34.
- Overbeek, J. Th. G., and Sparnaay, M. J., 1954, *Disc. Faraday Soc.*, **13**, 12.
- Parker, R. C., and Hatch, D., 1950, *Proc. Phys. Soc. (London)*, **B63**, 185.
- , and Marshall, P. M., 1948, *Proc. Inst. Mech. Engrs.*, **153**, 209.
- Pfefferkorn, G., 1953, *Naturwiss.*, **40**, 551.
- Rabinowicz, E., 1958, *Proc. Phys. Soc. (London)*, **B71**, 688.
- , and Tabor, D., 1951, *Proc. Roy. Soc. (London)*, **A208**, 455.
- Rehbinder, P., and Lichtman, V., 1957, *Proc. Second International Congress on Surface Activity* (London, Butterworth's Scientific Publications), Part III, pp. 563-580.

- Rubenstein, C., 1956, Proc. Phys. Soc. (London), *B69*, 921.
———, 1958, *Wear*, *2*, 85.
Savage, R. H., 1948, J. Appl. Phys., *19*, 1.
Schallamach, A., 1958, *Wear*, *1*, 384.
Scott, V. D., and Wilman, H., 1958, Proc. Roy. Soc. (London), *A247*, 353.
Semenoff, A. P., 1958, *Szvativanye Metallov* (Moscow, Mashgiz).
Sheppard, N., and Yates, D. J. G., 1956, Proc. Roy. Soc. (London), *A238*, 69.
Skinner, M. S., Savage, R. L., and Rutzler, J. E., 1953, J. Appl. Phys., *24*, 438;
1954, *ibid.*, *25*, 1055.
Sparnaay, M. J., 1957, in *Second International Congress of Surface Activity*
(London, Rutterworth's Scientific Publications), Part III, p. 442.
Tabor, D., 1955, Proc. Roy. Soc. (London), *A229*, 198.
———, 1957, *Wear*, *1*, 5.
———, 1959, Proc. Roy. Soc. (London), *A251*, 378.
Whitehead, J. R., 1950, Proc. Roy. Soc. (London), *A201*, 109.
Wilson, R. W., 1956, Proc. Roy. Soc. (London), *B68*, 625.

7. ELECTRONIC SURFACE PROPERTIES OF METALS AND SEMICONDUCTORS

The surfaces of all solids can be characterized by a region whose properties differ from those of the underlying material. In metals this region extends to a depth of approximately one atom layer; beyond this distance the interior is effectively screened from the surface by the high density of mobile electrons in the metal. In semiconductors the surface region is also of the order of one atom layer thick, but there is also a region of space charge directly below this surface layer because of the much lower density of mobile charge carriers. The presence of this space charge region and the existence of surface states impart to semiconductors electronic characteristics which are rather distinct from those of metals. Separate consideration will therefore be given to the electronic surface properties of these two classes of materials.

7.1. Electronic Surface Properties of Metals

The electronic surface properties of metals are determined by the interaction between the fields of the electrons in the surface region of the metal and electrons incident on this surface either externally or from within the metal itself. As in all collision problems, the results of this interaction are expressible in terms of transmission and reflection coefficients for electrons having a given momentum and energy. The main function of experiment is to provide sufficient data on the cross sections for these interactions so that the field can be properly reconstructed. Unfortunately, the incident beam in the usual surface scattering experiments is by no means monochromatic, but instead involves a wide

distribution over incident energies and directions; these distributions must be known sufficiently well before the experiments can be properly interpreted. Furthermore, the experiment itself rarely offers the most detailed information possible, yielding cross sections integrated over angles, or energy, or both.

That so much of a positive and definitive nature is known about these properties today is due to two rather fortunate circumstances, namely, that the incident flux at a given temperature can often be safely identified with the thermal flux of an electron gas in equilibrium at the same temperature, and that the integrated scattering effects are very small, once the conservation laws are satisfied. As a result, it has been possible to develop a reasonably straightforward theory for many of the observed effects.

7.1.1. The Experimental Situation

Somewhat disappointingly, experiment has so far yielded little detailed information on the surface fields. Aside from numerical values of the potential difference across the surface, the main conclusions which have emerged from experimental work are that the potential gradient is nowhere very sharp and that the classical image force is a reasonable representation of the force on electrons at all but the smallest distances from the surface.

Thermionic emission and field emission experiments have tested complementary aspects of this picture. In studies of thermionic emission, the flux is varied by controlling the temperature; in field emission experiments, the effective work function is altered by superimposing an external field on the surface fields. All of the important predictions of the Richardson, Schottky-effect, and Fowler-Nordheim equations describing these phenomena have been verified in detail. The supply current has been found to be independent of crystal orientation (Hutson, 1955), Maxwellian in thermionic emission (Shelton, 1957), Fermi-Dirac in field emission, and independent of the field (Müller, 1943). The emission current shows the proper dependence on the work function, the temperature, and the applied field over the entire range in which complicating effects such as space charge are negligible, provided, however, that the effective work function is defined by the combination of the applied fields and the classical image force. With this correction, the zero-field work function of the surface is the same in both thermionic and field emission phenomena.

The absolute magnitude of emission currents also agrees closely with theoretical prediction. In field emission, the major experimental limitation involves the determination of the surface area

and the local fields. A more accurate estimate of these quantities can be obtained in thermionic emission experiments, for which the agreement with the universal constant A ($= 120 \text{ amp/cm}^2 \cdot \text{K}^2$) is now within 15%.

It has become apparent that, of the great body of experimental information in this field, only very few and usually only the most recent experimental observations have any direct bearing on the successful correlation between theory and experiment. This is because the deceptively simple theory imposes the strictest experimental conditions, i.e., plane, atomically smooth, clean, single crystal surfaces subject to known, uniform, or at most slowly varying fields on the emitter, and, for successful detection, equally stringent requirements on the collector. It is not difficult to extend the original theory qualitatively to include a wide variety of more complicated experimental conditions (see, for example, the patch-effects discussed by Herring and Nichols (1949R)), and thus to make plausible the results of all too numerous experiments on ill-defined surfaces. However, the real test of a theory is its quantitative validity, and until recently, a number of detailed features of the basic theory were open to question.

A few of the most recent publications on field emission have emphasized both the need for accurate experimentation and the uncertainties inherent in much of the earlier published work. Müller, for example, has just reinterpreted some of his own older experiments and has concluded that earlier agreement of the velocity of field-emitted electrons with theory was accidental and due to the limited resolution of his instruments. His newest evidence (1959), interpreted by Young, (1959) indicates that, because of geometry, the retarding potential analyzer measures the total energy of emitted electrons, rather than their normal energy. Excellent agreement with the Fowler-Nordheim theory is still obtained, if a suitable correction is made for the collecting efficiency of the collector electrode; the absolute value of the observed currents is, however, not discussed.

Another crucial point, still not completely resolved, concerns the velocity distribution of thermionic electrons. Up to the time of Nottingham's (1956R) review article, the most reliable data indicated a serious deficiency of low-velocity electrons as compared to what would be expected from a Maxwellian distribution. Nottingham's preoccupation with this discrepancy pervades his whole article. Yet, since then, Shelton (1957) has achieved the ideal plane-parallel electrode configuration, using strong magnetic focussing so that small areas of single crystal faces confront each

other through an aperture. In this manner, Shelton has demonstrated a Maxwellian velocity distribution to high accuracy. The electrode arrangement permitted unambiguous absolute current density measurements, with the result that, on all surfaces examined, A was found to lie within 15% of the theoretical value. Detailed interpretation of these experiments and comparison with the earlier experiments is still in progress. Thus, Greenburg (1958) has just shown that the magnetic focussing field used by Shelton could not have influenced the observed velocity distribution. Nevertheless, taken at face value, the excellent agreement with theory found by Shelton seems to represent an important step forward in our understanding of thermionic emission. It implies that, with suitable experimental safeguards, attention may now be concentrated on the surface scattering problems *per se*.

One such problem, of fairly old standing, concerns the periodic deviations from the predicted Richardson-Schottky emission (see, for example, Good (1957) and Juenker (1957)). Qualitatively, these deviations result from the combined scattering of the surface potential proper and the Schottky hump in the image-force region. Quantitatively, however, agreement with the scattering predicted by all of the usual artificial surface potential models was poor, until an effort was made to reproduce more accurately the potential predicted by quantum mechanical calculation. In the recent work of Cutler and Gibbons (1958), for example, a successful correlation has been obtained with experiment, the gross features of a real surface potential having been retained in a calculable model. This agreement was achieved using a one-dimensional potential. It is quite possible that a three-dimensional potential may introduce effects of similar magnitude, but in this event pronounced changes in the phase of the periodic deviations on different crystals and different crystal faces would be expected.

Cutler and Gibbons (1958) also concluded that the reflection coefficient, R , in the absence of electric fields should have a value of about 0.4. This is an unusually high value, contrasting sharply with all earlier models which give, at most, values of 0.1 to 0.2. Because of the work of Shelton (1957), and of Fowler and Farnsworth (1958), among others, detailed experimental information on R is becoming available, both for average distributions of thermal electrons and for monochromatic beams. No direct comparison is possible because different surfaces are involved in the various experiments. It appears, however, that R is strongly energy-dependent at low energies, and approaches zero in this region.

Thus, a discrepancy still exists between theory and experiment, and among various experiments.

The most important conclusion which can be drawn at present is that experiments are now seriously approaching the stage where the fine structure of the interaction between electrons and an ideal metal surface can be studied systematically. So far only a small number of selected surfaces have been examined. A systematic investigation might establish whether or not R is dependent on crystallographic orientation of the surface—and provide possible evidence for the effect of band edges—as well as establishing the dependence of R on the direction of incidence of the electrons.

7.1.2. Theoretical Considerations

In anticipation of such refinements in experimental detail, corresponding improvements must be sought in the theory of surface fields. All models in present use are based upon a one-dimensional potential, except those which attempt to explain the variation in work function with crystal orientation. Although the change in work function with orientation has been understood qualitatively since the work of Smoluchowski (1941), a more reliable theory has not yet been developed, in good part because the uncertainty in the experimental information did not justify the effort involved. This situation is now on the verge of change. It must be kept in mind, however, that even the one-dimensional model has not been explored fully.

Bardeen (1936), for example, computed a self-consistent surface potential assuming a fixed exchange potential showing no fine structure. Recognizing the important role of the exchange potential in helping to account for electron correlation in the Hartree-Fock approach, as well as the fact that the surface potential is strongly determined by electron interactions, it will become necessary to carry out a more complete self-consistent calculation; perhaps it may even be advisable to go beyond the Hartree-Fock approximation in the direction of the many-electron approach. Calculations of this kind should also shed more light on the role of inelastic surface scattering.

The local potential and charge distributions in the surface must, of course, be derived from realistic electron wave functions compatible with this potential. However, unless the electron momentum lies close to a band edge, the reflection of incident electrons is mainly due to the surface fields rather than to the interior periodic potential, and hence the use of plane waves to compute reflection coefficients is adequate. In this connection, it is of

interest to note that one of the predictions of band theory (Herring and Nichols, 1949R), i.e., that the (111) face of Cu should show large reflection, was not substantiated by Fowler and Farnsworth (1958).

The same considerations apply to more atomistic models or theories. In addition, such theories must also account for the variation in work function with orientation, the temperature dependence of work functions, the dependence of R on tangential motion, the effects of lattice vibrations, and the surface energies of the crystal. On most of these topics, reliable experimental evidence is beginning to accumulate.

Photoelectric emission furnishes an additional probe into surface properties, because it samples the electron distribution at energies below the Fermi level. However, it is by far the most complex of the phenomena discussed here, and quantitative understanding of the experimental results is unsatisfactory in many respects. Qualitatively, an extension of the basic models to include the energy transfer of the photon—the so-called Fowler theory—reproduces adequately the gross features of the observed behavior near the emission threshold. Beyond that, the disagreement between theory and experiment must be traced to: (1) a relative dearth of modern experiments in this field, i.e., experiments carried out with the same care and with the same emphasis on fundamental details as are characteristic of recent thermionic and field emission experiments; (2) the theory must treat the difficult problem of an electron-photon collision in the presence of the surface fields; (3) the evidence points to at least two different mechanisms of photoelectron production, whose relative importance changes with the incident photon energy; (4) the band structure plays an important role but this is difficult to treat theoretically.

Nevertheless, photo-emission may eventually provide the most sensitive probe into surface properties. At present, however, convincing evidence is still lacking on such basic matters as whether or not the horizontal and vertical shifts in the origin of the Fowler curve with temperature can be assigned unambiguously to corresponding changes in the work function and reflection coefficient. Combined photo-emission and thermionic or field emission experiments on the same, and as nearly ideal surface as possible, seem to be essential in order to resolve this question. Another question concerns the detailed relations between optical and photoelectric properties, as, for instance, the effect of polarization of the incident light.

The interesting series of experiments by Thomas (1957),

Thomas and Mayer (1957) and Methfessel (1957) have suggested that yield curves can perhaps be understood quantitatively over a large range of incident energies by invoking only a single "volume effect" mechanism and taking into account the production of secondary electrons. This would remove many of the details of the photoelectric effect from the province of surface phenomena, and relate them more directly to the "bulk" mechanisms operative in secondary electron production. The papers mentioned also furnish a good review and critically perceptive summary of the current state of knowledge in the field, complementing the recent review by Weissler (1956R).

When properly understood, electron emission phenomena may also be used as analytical tools to determine the fine structure of the bulk electron properties of solids and how these are modified by phase transformations of various kinds. In fact, if done well, experiments of this type may even now provide considerable insight regarding fundamental questions. In this connection the experiments on Ni and Fe by Cardwell (1949, 1953) come to mind, both because of their bearing on the possibility of emission of polarized electrons (Dayhoff, 1959)—a problem which has been discussed extensively in the Russian literature (Vonsovski and Sokolov, 1951; Veksler, 1956)—and because of the information on the band structure of the transition metals which such experiments might disclose. At present, however, the experimental evidence is still ambiguous and the theories too primitive to draw any rigorous conclusions. Before this can be accomplished, we would require, on the one hand, detailed data on the energy and momentum distribution of emitted electrons, and, on the other, a better understanding of the effect of electron correlations and of the image force on the sharpness of definition of the band structure in the surface region.

All the above comments are made in full awareness of the tremendous practical or technical difficulties encountered in even the simplest surface experiments in which attempts are made to control the many factors that have a crucial bearing on final outcome. Nevertheless, to obtain the kind of basic information required, there is no alternative to careful experimentation. It may be hoped that recent advances in, for instance, ultra-high vacuum and electron detection techniques (sec. 8.5) may make such experiments possible in more than just the few laboratories currently equipped for exacting work of this nature.

7.2. Electrical Properties of the Surfaces of Semiconductors and Insulators

During the past decade, our understanding of the volume properties of semiconductors has been greatly augmented by a vast amount of research on single crystals. This has made it possible to attempt a study of single crystal surfaces by investigating their electrical properties. The surface of a semiconductor is fairly complex and must be considered as having extension in three dimensions rather than two. In the third dimension, the semiconductor surface may be roughly divided into the surface region and the space charge region directly beneath. The surface region may consist of only the first layer of atoms or perhaps the first few layers of atoms. If there is an oxide layer present, it may be considered part of this surface region. Just beneath the surface region there is usually a space charge zone which may vary in thickness from less than 100Å to half the width of the crystal. The energy level diagram for this region is similar to that of the bulk material. It can be represented by an almost empty conduction band and an almost filled valence band separated by an energy gap. In contrast to the region of space charge, the energy level diagram of the surface region is almost wholly unknown.

The space charge region arises from the trapping of excess charge in the surface region. If this is positive, the electrostatic potential energy in the region of space charge will be lowered for electrons and raised for holes. Thus, for an n-type semiconductor, electrons would accumulate in the space charge region and make the surface more conductive, while for a p-type semiconductor the positive charge in the surface region would repel holes; a depletion layer, i.e., a region which is less conductive than the bulk, would be formed as a consequence. In a completely analogous manner, if the net excess charge in the surface region is negative, the electrostatic potential energy in the space charge region will be raised for electrons and lowered for holes. Holes will accumulate in the space charge region of a p-type semiconductor while electrons would be repelled in an n-type semiconductor. In semiconductors, where both hole and electron conduction are possible, the carriers in the region of space charge may be either n- or p-type. For example, in germanium, an n-type crystal may have a p-type region of space charge due to a large amount of negative charge trapped in the surface region. This type of space charge region is called an inversion layer. If the net excess charge in the surface region is zero, no region of space charge

exists and the volume properties will extend right up to the surface region.

There are four possible states of an n-type semiconductor as the sign of the charge in the surface region changes from positive to negative:

(a) An n-type accumulation layer, i.e., an n-type space charge region on an n-type semiconductor;

(b) The flat band condition, there being zero excess charge in the surface region and no region of space charge;

(c) A depletion layer, which arises when the negative charge in the surface region is compensated by fixed donor atoms in the region of space charge;

(d) An inversion layer, i.e., a p-type region of space charge on an n-type semiconductor.

The great bulk of semiconductor surface research today is concerned with the characterization of the properties of the region of space charge, in order that the properties of the surface region may be deduced via the relations which connect the two. The most important of these is the condition of electrical neutrality, namely, the charge in the region of space charge must be equal and opposite to that in the surface region. After this brief introduction, we shall now review some aspects of our present knowledge about the electrical properties of various semiconductor surfaces and also some of the unsolved problems in the field. The main emphasis will be on germanium, since the surface properties of this material are best understood. It is hoped that its properties will serve as an example for the multitude of compound semiconductors which are of great industrial importance.

7.2.1. The Cleaned Germanium Surface

It has been shown by Farnsworth and his associates (Schlier and Farnsworth, 1959; Farnsworth et al., 1958) that a clean Ge surface can be produced and maintained in ultra-high vacuum. The surfaces were cleaned by ion sputtering followed by annealing at $\sim 600^\circ \text{C}$. Surfaces prepared in this manner, when investigated by low-energy electron diffraction methods, gave, to a very good first approximation, the diffraction maxima expected from the Ge surface lattice. With Si, this ion sputtering technique (and another method which only requires heating to just below the melting point) has also been shown to produce clean surfaces (Allen, Eisinger, Hagstrum and Law, to be published).

The work function and photoelectric threshold of a cleaned Ge surface have been measured by Dillon and Farnsworth (1957),

who found that these two quantities are approximately equal. This would be normal for a metal; however, for a semiconductor such as Ge, it is not. The photoelectric threshold is the difference in energy between the first occupied states and the vacuum level. In semiconductors the first occupied level of any consequence is the valence band. The work function is defined as the difference between the Fermi level and the vacuum level. Thus, for a semiconductor in the flat band condition, the difference between the photoelectric threshold and the work function should be $E_F - E_v$, which, for germanium, is of the order of 0.3 eV; however, experiment gave a value of almost zero. To explain this result, it must be assumed that either the valence band edge has been raised at the surface so that it is coincident with the Fermi level, or the conduction band has been lowered so that it is coincident with the Fermi level. The first situation would correspond to a p-type and the second to an n-type region of space charge.

The problem was resolved by the data obtained from field-induced surface conductivity measurements (Handler, 1957) on the cleaned germanium surface. These data show that the dominant carriers in the region of space charge are holes and therefore the valence rather than the conduction band is degenerate near the surface. Measurements of the conductivity and Hall coefficient (Missman and Handler, 1959) of these holes showed that their number was large and that the valence band must be close to degenerate at the surface in order to accommodate them. Thus, two different types of experiments, the photoelectric threshold experiment, where an electron is ejected through the surface region, and the surface conductivity experiments, where the transport properties of electrons moving parallel to the surface are measured, both give a similar picture of the germanium surface. The surface conductivity data indicate that there are approximately 10^{12} holes per cm^2 in the region of space charge. Since the surface is electrically neutral, there must be an equal number of electrons in the surface region. It is assumed that these electrons are in localized surface states with energies in the forbidden band. The exact nature of these states is, however, unknown at present.

From an atomic viewpoint, it can be shown that these acceptor-like surface states are probably associated with the unfilled orbitals of the germanium surface atoms. On a (111) germanium surface, the distance between surface neighbors is 1.63 times the nearest-neighbor distance in the bulk. Therefore, to a first approximation, the surface atoms can be considered as atoms at the

end of a large molecule. These surface atoms form three covalent bonds with the atoms in the crystal below and have one free orbital above the surface plane. This free orbital is occupied by one electron and may accept a second electron from the bulk. In so doing, the surface atom achieves a stable octet of eight electrons and becomes negatively charged. A hole is generated nearby and this hole is free to move, as is observed experimentally.

7.2.2. The Effect of Gases on Semiconductor Surfaces

The effect of gases on the oxide-covered germanium surface has been extensively studied over the past few years. The oxide-covered surface is usually produced by means of an aqueous etch; it is not stable in air, but changes rapidly during the first day and somewhat more slowly thereafter. The structure and chemical composition of the superficial oxide layer are unknown and its thickness may vary from only a few Å up to 100Å. Moreover, the properties of such oxide-covered surfaces are somewhat different from those of an atomically clean surface which has been exposed to pure oxygen. Nonetheless, the results obtained have added greatly to our understanding of surfaces.

For example, it has been shown that the addition of water vapor, or other electron donor-type molecules to the etched surface can make the region of space charge more n-type, while the addition of oxygen or other electron acceptor-type molecules can make it more p-type. Using these techniques, Bardeen and Brattain (1953) showed very clearly the existence of the space charge region. Since the early work of Bardeen and Brattain, those properties which may be derived from a solution of the one-dimensional Poisson equation have been substantiated experimentally (Garrett and Brattain, 1955; Kingston, 1956, 1957R). These later experiments have shown that the surface region of etched Ge exhibits a number of discrete surface states in the vicinity of the energy gap. The densities of these surface states are of the order of 10^{11} per cm^2 . Some states appear to be effective recombination centers while others do not. Capture cross sections for some states are estimated to be as high as 10^{-14} cm^2 (Rupprecht, 1958). These are in fast equilibrium with the region of space charge, their relaxation times being of the order of microseconds or less at room temperature. At the present time, no particular structure or defect has been associated with these surface states, but they are assumed to lie at or very near the semiconductor oxide layer interface. Similar states are also found on etched Si surfaces (Statz, deMars, Davis and Adams, 1957).

A second group of states is observed on etched Ge and Si surfaces which differ from the fast states mentioned in the preceding paragraph. These are the so-called slow states. Their densities are much higher and their relaxation times are of the order of seconds to hours. The origin of these slow states is not clear. One hypothesis suggests that they are associated with adsorbed gases, while another suggests that they are intrinsic to the oxide. If a thick oxide layer is grown on the surface of Si or Ge, the relaxation effects associated with the slow states seem to disappear, either indicating that the relaxation times have increased by orders of magnitude, or that the states have been removed (Lasser, Wysocki and Bernstein, 1957). The addition of water vapor will reduce the relaxation times of these thick oxide layers somewhat but will not return them to the values observed for the etched surface.

In compound semiconductors such as ZnO, a space charge region may also be induced by departures from stoichiometry. Thomas and Lander (1957) have shown that adsorbed zinc atoms act as donors at the surface, giving up electrons to the crystal to form an n-type region of space charge. Introduction of oxygen to the surface removes the n-type region and may, under some conditions, induce a depletion layer. Heiland (1957) has shown that atomic hydrogen will produce an n-type region of space charge on zinc oxide which may be removed by the introduction of oxygen. Thus, we see that the state of a zinc oxide surface depends very strongly on the partial pressures of the ambient gases. Collins and Thomas (1958) have shown that an n-type space charge region may be induced in zinc oxide by illumination with light of energy equal to the fundamental optical adsorption edge. In this process, either neutral oxygen atoms are formed at the surface or oxygen is desorbed.

In compound semiconductors, ionic mobility under the influence of the high fields in the space charge region may be sufficient to cause ion migration. This effect is of great importance in the initial stages of oxidation, and represents an attempt of the system to lower its electrostatic energy.

A vast amount of literature has been published on the surface properties of many other compound semiconductors. In principle, the electronic surface properties of these materials should be explainable in similar terms, once the bulk properties are sufficiently well known. In most cases, however, such information is lacking, and the materials themselves are not pure and homogeneous enough for so detailed an analysis.

The same applies to insulators. Here, however, non-reproducible and slowly time-dependent space charge distributions make the analysis even more difficult. Thus, although static electrification (triboelectricity) is of industrial importance and represents a problem of great scientific antiquity, the fundamental processes involved are still not well understood (Rose and Ward, 1957; Wagner, 1956; see also Symposium on Static Electrification, 1953R).

7.2.3. The Mobility of Charge Carriers in the Region of Space Charge

Electrons and holes moving in the region of space charge are constrained to move in a narrow potential well. A characteristic of these wells is that they become narrower as they become deeper. Thus, with increasing depth, the carriers will have a greater probability of colliding with the surface. If the process of collision with the surface acts to scatter the charge carriers nonspecularly, then the mobility will be reduced below the value it has in the bulk. The theory of this effect has been worked out by Schrieffer (1955) and it has been experimentally verified for shallow potential wells by Zemel and Petritz (1958). For very deep wells, most of the carriers are moving in a channel less than 100Å wide. In this case, the assumption that the region of space charge has an energy level diagram similar to that of the bulk is no longer valid.

7.2.4. Barrier Limited Recombination at the Germanium Surface

One other problem which arises in dealing with deep space charge regions, is the role which the space charge barrier plays in limiting the recombination of minority carriers at the surface. In the papers of Bardeen and Brattain (1953), Stevenson and Keyes (1955), Many and Gerlich (1957) and Wang and Wallis (1959), it is explicitly assumed, in adapting the Shockley-Read (1952) model of recombination via traps to the surface, that the volume and surface portions of the conduction bands are in good thermodynamic equilibrium. Brattain and Garrett (1955) have shown, however, that the barrier may limit the recombination rate under conditions where the region of space charge is much wider than the mean free path of the carriers. The barrier for electrons on a cleaned germanium surface is 0.3 eV and according to the criterion of Brattain and Garrett it may become rate limiting. The data of Law and Garrett (1956) on the surface recombination velocity of cleaned germanium seem to support this view. Law and Garrett find that the surface recombination velocity is low

as long as the barrier is of the order of 0.3 ev. As soon as the barrier begins to fall below this value the surface recombination velocity begins to increase quite rapidly. These results indicate that the Shockley-Read model of recombination may not be correct when large space charge barriers are present.

7.2.5. The Surface Region

It should be apparent, from the previous discussion, that our knowledge and understanding of the properties of the space charge region has increased considerably, especially in the last few years. However, our knowledge of the surface region is still poor. The main difficulty is that it has generally not been possible to investigate this region directly. Most of our information concerning the surface region has come from indirect measurements of the properties of the space charge region or of the properties of the gas in contact with the solid surface. The state of the surface region can only be inferred from measurements of these kinds. Moreover, many different states of the surface region will give rise to identical space charge regions. The advantage of semiconductor surface studies is that the data corresponding to the space charge region may be correlated with the data obtained for the gas phase so as to delineate the actual state of the surface region.

What is needed at the present time are experimental techniques capable of measuring the properties of atoms which are physically a part of the surface region. One such method, namely infrared adsorption spectroscopy of chemisorbed molecules, has recently been developed by Eischens (1958) and promises to provide valuable insight into the properties of the surface region. Other methods which are potentially capable of giving information about the surface atoms themselves include electron paramagnetic resonance and nuclear magnetic and quadrupole resonance techniques. Unfortunately, great difficulties arise in applying these techniques to surface atoms. In the case of electron paramagnetic resonance, a system must be found which not only has a large density of unpaired electrons at the surface, but also a suitable relaxation time. In the nuclear magnetic and quadrupole resonance techniques, the relaxation problems are not as severe, but the sensitivity is somewhat poorer. While the electron paramagnetic resonance technique is capable of sensing 10^{11} unpaired electron spins for a line width of one gauss, nuclear techniques would require 10^{18} or more surface nuclei. Consequently, the application of nuclear magnetic or quadrupole resonance techniques

would seem to require the use of extremely fine powder samples. By employing nuclear quadrupole resonance techniques, however, a value could be obtained for the electric field gradient about the surface nuclei; this would provide direct information about the electron density at the surface. Methods of this kind hold great promise for the future and it is hoped that the next few years will see increased activity in many of these areas.

7.3. References

7.3.1. Reviews and General Articles

- Handler, P., 1957, in *Semiconductor Surface Physics*, R. H. Kingston, editor (Philadelphia, University of Pennsylvania Press), pp. 23-51.
 Herring, C., and Nichols, M. H., 1949, *Rev. Mod. Phys.*, **21**, 185.
 Kingston, R. H., editor, 1957, *Semiconductor Surface Physics* (Philadelphia, University of Pennsylvania Press).
 Nottingham, W. B., 1956, *Handbuch der Physik*, S. Flügge, editor (Berlin, Springer-Verlag, Vol. XXI, pp. 1-175).
 Symposium on Static Electrification, 1953, *Brit. J. Appl. Phys.*, Suppl. 2.
 Weissler, G. L., 1956, *Handbuch der Physik*, S. Flügge, editor (Berlin, Springer-Verlag), Vol. XXI, pp. 304-382.

7.3.2 Original Articles

- Allen, F. G., Eisinger, J., Hagstrum, H. D., and Law, J. T., to be published.
 Bardeen, J., 1936, *Phys. Rev.*, **49**, 653.
 ———, and Brattain, W., 1953, *Bell Syst. Tech. J.*, **32**, 1.
 Brattain, W. H., and Garrett, C. G. B., 1955, *Bell Syst. Tech. J.*, **35**, 1019.
 Cardwell, A. B., 1949, *Phys. Rev.*, **76**, 125; 1953, *ibid.*, **92**, 534.
 Collins, R. J., and Thomas, D. G., 1958, *Phys. Rev.*, **112**, 888.
 Cutler, P. H., and Gibbons, J. J., 1958, *Phys. Rev.*, **111**, 394.
 Dayhoff, E. S., 1959, *J. Appl. Phys.*, **30**, 234S.
 Dillon, J., and Farnsworth, H. E., 1957, *J. Appl. Phys.*, **28**, 174.
 Eischens, R. P., 1958, *J. Chem. Ed.*, **35**, 385.
 Farnsworth, H. E., Schlier, R. E., George, T. H., and Burger, R. M., 1958, *J. Appl. Phys.*, **29**, 1150.
 Fowler, H. A., and Farnsworth, H. E., 1958, *Phys. Rev.*, **111**, 108.
 Garrett, C. G. B., and Brattain, W., 1955, *Phys. Rev.*, **99**, 376.
 Good, R. H., Jr., 1957, *J. Appl. Phys.*, **28**, 1405.
 Greenburg, J., 1958, *Phys. Rev.*, **112**, 1898.
 Heiland, G., 1957, *Z. Physik*, **148**, 15; 1957, *ibid.*, **148**, 28.
 Herring, C., and Nichols, M. H., 1949, *Rev. Mod. Phys.*, **21**, 1.
 Hutson, A. R., 1955, *Phys. Rev.*, **98**, 889.
 Juenker, D. W., 1957, *J. Appl. Phys.*, **28**, 1398.
 Kingston, R. H., 1956, *J. Appl. Phys.*, **27**, 101.
 Lasser, M., Wysocki, C., and Bernstein, B., 1957, *Phys. Rev.*, **105**, 491.
 Law, J. T., and Garrett, C. G. B., 1956, *J. Appl. Phys.*, **27**, 656.
 Many, A., and Gerlich, D., 1957, *Phys. Rev.*, **107**, 404.
 Missman, R., and Handler, P., 1959, *J. Phys. Chem. Solids*, **8**, 109.
 Methfessel, S., 1957, *Z. Physik*, **147**, 442.

- Müller, E. W., 1943, *Z. Physik*, **130**, 261.
———, 1959, *Phys. Rev.*, **113**, 110.
Rose, G. S., and Ward, S. G., 1957, *Brit. J. Appl. Phys.*, **8**, 121.
Rupprecht, G., 1958, *Phys. Rev.*, **111**, 75.
Schlier, R. E., and Farnsworth, H. E., 1959, *J. Appl. Phys.*, **30**, 917.
Schrieffer, J. R., 1955, *Phys. Rev.*, **97**, 641.
Shelton, H., 1957, *Phys. Rev.*, **107**, 1553.
Shockley, W., and Read, W. T., 1952, *Phys. Rev.*, **87**, 835.
Smoluchowski, R., 1941, *Phys. Rev.*, **60**, 661.
Statz, H., deMars, G., Davis, L., Jr., and Adams, A., 1957, *Phys. Rev.*, **106**, 455.
Stevenson, D. T., and Keyes, R. J., 1955, *Physica*, **20**, 1041.
Thomas, D. G., 1957, *Z. Physik*, **147**, 395.
———, and Lander, J. J., 1957, *J. Phys. Chem. Solids*, **8**, 318.
———, and Mayer, H., 1957, *Z. Physik*, **147**, 419.
Veksler, A. Z., 1956, *JETP*, **2**, 266.
Vonsovski, S. W., and Sokolov, A. V., 1951, *Doklady. Akad. Nauk. S.S.S.R.*, **76**, 197.
Wagner, P. W., 1956, *J. Appl. Phys.*, **27**, 1300.
Wang, S., and Wallis, G., 1959, *J. Appl. Phys.*, **30**, 285.
Young, R. D., 1959, *Phys. Rev.*, **113**, 115.
Zemel, J., and Petritz, R., 1958, *Phys. Rev.*, **110**, 1263.

8. INTERACTIONS OF SURFACES WITH GASES

This section deals with adsorption, catalysis and other surface reactions, particularly gas-solid reactions such as oxidation. A short discussion concerning impact phenomena is also included as well as a brief description of modern experimental techniques. Adsorption plays a basic role in many surface phenomena. It constitutes the primary step in catalysis, oxidation, and countless other gas-solid reactions. We therefore begin our discussion of the interactions of surfaces with gases by considering the adsorption of gases on solids.

8.1. Gas-Solid Adsorption

There are two general types of adsorption on solids. The first, physical (or physisorption), is highly non-specific and occurs on all solids with all gases. The forces involved are of the van der Waals type, i.e., fairly long range but only weakly attractive. Physisorption is appreciable only in a temperature-pressure range near the condensation point and many layers may usually be formed reversibly.

The second type of adsorption, i.e., chemisorption, is a much more specific phenomenon. Energies of adsorption are high, from one to several eV, and this tight binding is confined to one or at most two atom layers, which may be stable even at high temperatures and extremely low pressures. The magnitude of the energies

(and forces) associated with chemisorption indicates that electronic rearrangement is involved. Often, this can be accomplished only at the cost of severe modification or outright dissociation of the adsorbate. Frequently, there is also an overlap between chemisorption and true surface reaction involving irreversible rearrangements of the substrate.¹ The irreversibility of chemisorption is due both to the high energies involved and to actual surface reaction.

Chemisorption may also occur as a slow, activated process, particularly on semiconductors and insulators. On clean metals, however, it is generally rapid, even when dissociation occurs, as in the adsorption of H_2 or O_2 on tungsten.

With this brief introduction, these two general types of adsorption will now be discussed separately, although it is clear that some overlap is inevitable.

8.1.1. Physisorption

The existence of multilayer adsorption was first postulated by Brunauer and Emmett who noted that many adsorption isotherms exhibited inflections at pressures of about $0.1 p_0$, with p_0 being the condensation pressure. Beyond this point a slow rise occurred in the amount adsorbed, followed by a steep increase near p_0 . The initial inflection (point B) was presumed to correspond to the completion of a monolayer, this being followed by multilayer adsorption and finally, near p_0 , by condensation of liquid. Brunauer, Emmett and Teller (1938) incorporated these ideas in the so-called BET-theory of adsorption which leads to a mathematically simple isotherm. However, as first pointed out by Halsey (1948), the model on which the theory is based is highly idealized. Basically, the assumptions of the model are equivalent to assuming simultaneous partial formation of many layers. Since it is physically unreasonable to expect the stacking of isolated molecules on top of each other, multilayer formation on a homogeneous surface should proceed in steps, a given layer having a finite population only after the previous layer has been virtually completed. Step-wise adsorption of this kind has, in fact, been observed for inert gases on tungsten field emitters (Gomer, 1958). On most adsorbents, however, a wide variety of crystal surfaces is exposed and there is also a great deal of pore structure and capillarity at points of contact between particles. This leads to a

¹ Rearrangement resulting from adsorption can be investigated, at least for moderately high melting metals, by field ion microscopy, although little work has been carried out in this direction so far.

smearing out of the idealized step isotherms and results in the quasi-BET behavior observed for many real systems.

Although the theoretical inadequacies of the BET model are apparent, the identification of the inflection point, B, with the completion of the first layer is often qualitatively correct. This is because the heat of adsorption associated with this layer is always somewhat larger than that for subsequent layers, often sufficiently so that the effect is not masked by the heterogeneity of the adsorbent. In any event, point B can legitimately be identified with a change in adsorption energy. This fact lends enormous practical utility to the BET isotherm as a means of at least estimating the real surface area of substrates for chemisorption and catalysis, since a size which is only slightly arbitrary can be assigned to the adsorbate molecules.

The correct formulation of the statistico-mechanical equations for physisorption is not difficult but the solution has so far proved impossible without drastic simplifications. The liquid slab model of Halsey, Hill and Frenkel (Hill, 1953R) aims at accomplishing this by treating physisorption as being equivalent to the formation of a liquid-like slab on the surface. Isotherms of the type

$$\ln (p/p_0) = -k/V^r \quad (8.1)$$

are obtained, which seem to be obeyed experimentally near p_0 with $r \sim 3$, V being the volume adsorbed and k a constant. Nevertheless, finding a realistic solution of the statistic-mechanical equations for physisorption must still be regarded as the major unresolved theoretical problem in this field. The magnitude and nature of the forces is fairly well understood in many though not all cases.

On the experimental side, a great need exists for studying homogeneous, or at least completely characterized substrates, in order to avoid smearing out steps in the adsorption isotherms. Recent studies by Ehrlich et al. (1958a) and by Gomer (1958) utilizing field emitters as adsorbents seem to have disclosed at least one promising line of attack to this problem.

Among other things, studies using field emitters as substrates have shown that multilayers may remain liquid-like considerably below their bulk melting point (Gomer, 1958). This may perhaps be connected with the inhibition by the substrate lattice of long-range order in the adsorbate. Before leaving the subject, another recent but only partially understood observation will be noted. It was first reported by Mignolet (1955), and subsequently confirmed by field emission experiments (Gomer, 1958; Ehrlich et al., 1958), that physisorption of inert gases and of molecules on

metals produces a large dipole moment (0.1 – 0.8 Debye) in the adsorbate, the negative end of the dipole pointing toward the surface. Despite this, however, the adsorption energies appear to be "normal." It is interesting to speculate on the mechanism of dipole formation (charge transfer, polarization?) and on the reasons for the lack of severe anomalies in the resultant heats of adsorption.

8.1.2. Chemisorption

The specificity encountered in chemisorption makes the use of clean and, if possible, homogeneous or at least crystallographically well-characterized surfaces an absolute necessity if any significance is to be attached to the experimental results. Once a clean surface has been attained, its future state depends on the pressure of gaseous impurities in the system. If a sticking coefficient of 0.1 (the order of magnitude indicated by most recent work) is assumed, a monolayer is formed in 1 second at a pressure of 10^{-6} mm at 300°K. Working vacua of the order of 10^{-9} mm or even better are therefore essential. These are generally realizable for metals but not for some technically important oxides and sulfides, whose vapor or dissociation pressures limit the attainable vacuum. Unfortunately, the great bulk of previous experimental work in this field serves no more useful purpose than to illustrate this dolorous lesson over and over again.

Notable exceptions may be found as far back as thirty years ago when Langmuir (1912) laid the foundations for our present understanding of chemisorption phenomena. Aside from the work of Langmuir, that of Roberts (1935, 1937) on wires and of Beck and his associates (Wheeler, 1953R) on films also seems to have substantially met both of the requirements mentioned above. Electric heating of wires or filaments at temperatures near the melting point, such as employed by Langmuir and Roberts, is frequently found to produce a clean surface, while evaporated films often getter their own vacuum. In recent years, the popularization of ultra-high vacuum techniques (Alpert, 1953; Alpert and Buritz, 1954; Pollard, 1959R) and the development of other new tools (sec. 8.5) have more or less revolutionized the study of chemisorption, at least on metals; stated more accurately perhaps, the availability of these new techniques has reopened the paths so clearly indicated by the best of the early work in the field.

8.1.3. Chemisorption on Metals

From the early work of Langmuir, Roberts and Beeck, together with more recent investigations, a reasonably consistent picture of chemisorption on metals has now emerged. Although significant advances in our understanding have been made, it should be emphasized that many important questions are yet to be answered or resolved.

Adsorption is rapid and occurs with almost zero activation energy on most metals, despite the common occurrence of dissociation (Wortman et al., 1957; Gomer, 1959R, 1961R; Becker, 1955R; Hickmott and Ehrlich, 1958; Farnsworth and Woodcock, 1957R). Sticking coefficients are high—of the order of 0.1 to 0.5—and vary with coverage (Becker, 1955R; Ehrlich, 1956; Hickmott and Ehrlich, 1958). It seems probable that the sticking coefficients differ from unity because re-evaporation of physically preadsorbed molecules can precede chemisorptive binding (Hickmott and Ehrlich, 1958). Since there seem to be definite, crystallographically determined sites for adsorption (Wheeler, 1953R; Gomer, 1959R; 1961R; Farnsworth and Woodcock, 1957R), the search that a molecule makes for a favorable site (or when dissociation is involved, an adjacent pair of ad-sites) will be shorter and hence more successful at low surface coverages. Beyond a certain coverage, there may also be a real scarcity of adjacent pair sites, as first pointed out by Roberts (1937), so that some molecular chemisorption may occur even with molecules which normally dissociate when chemisorbed.

Gomer and coworkers have found (Gomer et al., 1957; Gomer and Hulm, 1957; Gomer, 1959R) that approximately 80 percent of the chemisorbed layer can be formed with essentially zero activation energy for hydrogen or oxygen on W or Ni at temperatures between 4° and 20°K. For hydrogen at high coverages, a reversible equilibrium seems to occur between the molecularly and atomically adsorbed fractions. True activation energies of dissociation on clean surfaces seem to be observed only with CO, where the activation energy is so high that dissociation does not occur at all, and with N₂, where evidence for physisorption has been obtained at 77°K and chemisorption at 300°K (Ehrlich et al., 1958b). The reason for this is most probably that the intersection of the potential curves for pure chemisorption and pure physisorption occurs at a point where physisorption is still attractive and chemisorption is already so; consequently, splitting of the levels results in a smoothly attractive ground state potential.

Energies of adsorption decrease somewhat with increasing

surface coverage. This is reflected by correspondingly larger reductions in the heats of adsorption of dissociating molecules (Wheeler, 1953R). Recent mobility studies (Gomer, 1959R) seem to explain this behavior, at least in part. It is found that diffusion occurs with activation energies ranging from about 10 percent to greater than 45 percent of the adsorption energies and that this ratio, as well as the absolute magnitudes, depends upon the atomic surface configuration. The values increase in both cases with increasingly open surface structure. It is also found that sites with inherently different binding energies occur on one and the same crystal face, except perhaps on the most closely packed surfaces, as an unavoidable consequence of the discrete atomic structure of the metal. This strongly suggests that the observed decrease in the heats of adsorption results from a built-in heterogeneity, at least at moderate coverages. At high coverages, ad-ad particle interactions (dipole-dipole repulsion and other electronic effects (Boudart, 1952; Gomer, 1953)) can become important and may account, in part, for the atomic-molecular switching already noted.

Relatively little work of a reliable nature has been carried out on desorption, with the exception of flash filament experiments of Ehrlich (1956) and Hickmott and Ehrlich (1957). Detailed kinetic data are almost completely lacking. Mobility usually precedes desorption (Gomer, 1959R) so that only average rates are obtained over various faces for microcrystalline substrates. These averages weight the most strongly bound sites rather heavily. Generally, atomic desorption (in the case of polyatomic stable end products) requires much larger energies than molecular desorption and therefore occurs only at much higher temperatures. Thus, H_2 evaporates rapidly from W at 600° to 700°K (Gomer et al., 1957; Gomer, 1959R), while atomic desorption is inappreciable below 1000°K (Langmuir, 1921).

It is necessary to consider the electronic states of ad-particles. Contact potential measurements indicate that the binding is usually covalent rather than ionic although moderately large bond dipole moments are common. Attempts to correlate these (or the contact potential data directly) with heats of binding have been made (Ely, 1950; Stevenson, 1955) but must be viewed critically because: a) much of the older data are unreliable or subject to considerable error; b) there is no single characteristic heat of adsorption for a given system.

The criterion for covalent or ionic binding may be taken as the sign of the inequality.

$$H_a \geq (\phi - I)/K + H_{ion} \quad (8.2)$$

where H_a and H_{ion} are the "intrinsic" adsorption energies of atom and ion in volts, ϕ and I the work function and ionization potential, K the dielectric constant of the substrate. On metals, the small effective dielectric constant results in polarized but covalent adsorption for most substances of high vacuum ionization potential. As pointed out later, field desorption (Gomer, 1959) is sometimes able to distinguish experimentally between ionic and covalent binding, showing, for instance, that Th or Ba on W may exist as ions at least at low coverages.

While the direction of electron transfer in ad-bonding is usually indicated by the sign of the contact potential (when correctly obtained) other methods have also been used. Thus, changes in paramagnetism (Selwood, 1957R) or electrical conductivity (Suhrmann, 1955R) resulting from adsorption have been employed for this purpose. Unfortunately, these methods require the use of evaporated films, which are microcrystalline and sometimes almost amorphous, so that a serious question arises concerning the validity of the interpretations even if the experimental conditions are adequate. It would be interesting to measure conductivity changes produced by adsorption on a monocrystalline metal whisker of very small volume to surface ratio.

The detailed nature of the ad-bond in chemisorption is poorly understood. Only a few highly simplified quantum mechanical calculations have been carried out (Huang and Wylie, 1950). In view of the lack of detailed knowledge concerning the structure and properties of clean metal surfaces, this is hardly surprising. Although we are quite unable to make *a priori* energy or dipole moment calculations (even on the semi-empirical level of molecular spectroscopy) certain qualitative statements can be made (Baker and Jenkins, 1955R). For example, it appears that strong adsorption will occur only if substrate and adsorbate, respectively, have available electrons and orbitals. Transition metals fit particularly well the roles of both donor or acceptor because of their partially filled *d*-bands (Trapnell, 1954R); these bands are relatively narrow and correspond to fairly localized electrons. The *d*-electrons have a tendency to lower their energies by bonding, which enables them to approach the bonding nucleus, thereby giving rise to a decrease in potential energy. This is true only of *d*-electrons in partially filled bands. As filling nears completion, more overlap and exchange occurs within the band and hence less need arises for outside alliances.

Ad-bonds may be regarded as localized impurity states which

lie somewhere below the Fermi level. Demotion of an electron to this state is energetically most profitable if the electron comes from the vicinity of the Fermi level. An energetically favorable situation exists for *d*-electrons in partly filled bands, but not for electrons in filled bands. The *s*-electrons form a broad, non-localized bands and therefore have relatively little to gain energetically from localization by bonding.

The effect of surface structure, so intuitively obvious in terms of discrete bonds or nearest-neighbor interactions, can also be justified quantum mechanically on the basis of maximum sampling of Coulomb interactions. Any explanation in terms of resonance really provides only an equivalent description.

This rather qualitative picture seems to explain the observed behavior (Wheeler, 1953R) fairly well. A word of caution must be added, however. Many experimental heats of adsorption are suspect because of doubts concerning the initial cleanliness of the surface, and because of the fact that surface reactions can occur along with adsorption.

8.1.4. Chemisorption on Semiconductors

Electron sharing or exchange also plays a decisive role in chemisorption on semiconductors, as previously indicated in the discussion of the electronic properties of semiconductors (sec. 7.2). Early proof of its importance was provided by Wagner and Hauffe (1938), by Garner (1949), and others (Hauffe, 1955R). At present, it is believed by many workers that the end result of chemisorption on semiconductors is the formation of adsorbed ions. This is certainly plausible on the basis of Eq. (8.2) since the dielectric constant of semiconductors is high. The localization of charge at the surface leads to a space charge region of opposite sign, as discussed in detail in section 7.2.1. Ionic adsorption therefore rapidly becomes activated since the charge carriers diffusing to the surface must overcome the resultant potential barrier. At the same time, the electrostatic potential resulting from the diffuse double layer raises the ad-levels, that is, decreases the heat of adsorption, until, at equilibrium, the initial heat of adsorption at zero coverage is just balanced by the electrostatic decrease. This generally occurs when the surface coverage θ , is of the order of 0.01.

Electrical measurements of various kinds have confirmed this general picture. However, the details are still incomplete. Thus, we do not know at present what role is played by covalent bonding on semiconductors (Sandomirski, 1957) at coverages where ionic

adsorption no longer occurs. It is quite probable that this type of bonding would influence the electrical conductivity, by localizing charge carriers, almost as strongly as pure ionic adsorption. Covalent adsorption might also be activated, since there is some possibility that it would involve the "opening up" of surface bonds between the substrate atoms.

At present, adsorption on semiconductors is being investigated by quite indirect means. Consequently, very little is known as yet about the detailed kinetics of adsorption or desorption. For example, we do not know the steps by which a diatomic molecule becomes dissociated and ionized on the surface. In fact, relatively little is known about the nature of the ad-species themselves.

Similarly, little information is available on the relations between atomic surface structure and ionic or covalent adsorption. It is likely that image potential effects are important in ionic adsorption and that resonance effects are of significance in covalent adsorption (as with metals). Since these both depend on structure, the atomic configuration at the surface should be influential. For the same reason, structural imperfections may be expected to play an important role (Gray and Savage, 1959).

Another phenomenon may occur on semiconductors which complicates both the execution of experiments and their interpretation. The presence of the space charge region with its high local electrostatic fields leads to a certain amount of field transport of ions. In a sense, field transport represents the precursor step for surface reactions, and it may produce substantial changes in the structure and electronic properties near the surface. Isotope exchange experiments indicate that substrate atoms can participate in chemisorption as well as in surface reactions proper. It might be pointed out in this connection that the chemisorption of H_2 on NiO is accompanied by reaction even at $50^\circ C$ (Schlosser and Haufler, unpublished), as in the adsorption of CO on NiO at $100^\circ C$ (Gossel and Haufler, unpublished). In other systems, e.g., CO_2 on NiO, surface reactions can probably be neglected at ordinary temperatures. These effects are undoubtedly the result of field transport. The occurrence or absence of secondary reactions of this kind must depend on the diffusion potential at the surface, which may be negligibly small if both electron donating and accepting processes occur (as in some catalytic reactions) with comparable rates.

There are experimental indications that such complications may occur. If they do, adsorption will be irreversible and the desorption product will differ from the initial adsorbate. Unfortunately, at

the temperatures required for desorption, reaction frequently does take place even if the initial adsorption was not accompanied by reaction. If "adsorption" can be shown to be purely ionic at coverages in excess of 0.01 this would also indicate reaction.

The above discussion points up some of the major unsolved problems in adsorption on semiconductors. In a way, these represent merely a restatement, with shifted emphasis, of the problems already mentioned in connection with the electronic surface properties of these materials. It is unlikely that significant theoretical advances will be made until a firm background of experimental information is developed. However, before much effort is placed in devising new experiments, improvement in the conditions under which such experiments are carried out is urgently needed. First and foremost, better knowledge and control of the surface structure will be required. In particular, the following characteristics of the substrate must be known in detail: 1) the total surface area; 2) the crystallographic orientation and atomic configuration at the surface, and 3) the electronic structure, i.e., the Fermi level, energies of band edges, presence, number and type of surface states, etc. (sec. 7.2).

Only the first of these conditions can normally be fulfilled with relative ease. At present, it seems improbable that condition (2) has been satisfied in any experiment with surfaces other than Ge (or Si?), for which it is possible, at least theoretically, to attain an ultimate state of perfection. Great difficulties arise, for example, with compound semiconductors, whose surfaces are profoundly influenced by ambients and fundamentally altered by attempts at cleaning. It is, therefore, necessary to examine each working surface critically (by electrical and/or diffraction methods) and to place less reliance on the theoretical structure than is permissible with surfaces which can really be cleaned. Compliance with condition (3) depends, in turn, on how successfully the atomic configuration at the surface can be characterized and controlled, since the initial state of the surface has a marked affect on its electronic properties.

Assuming that efforts in these directions will prove fruitful, new and diverse experiments are required. For instance, it will be apparent from what has been said about the uncertainties connected with the adsorbed state, that simultaneous knowledge of the amount adsorbed and electrical or magnetic changes produced by adsorption is needed. It is also probable that experiments on photo-desorption or photo-adsorption, properly combined with electrical and volumetric studies, will prove extremely useful in

elucidating the nature of the electronic levels and carriers involved in adsorption. At the moment work of this kind is in a fairly embryonic state (Terenin and Solonitzin, 1959; Medved, 1958).

8.2. Heterogeneous Catalysis

Catalysts do not affect chemical equilibria but can increase the rate at which equilibrium is attained. For reactions which occur on surfaces, this enhancement of the rate of approach to equilibrium results from a decrease in the activation energy of one or more steps in the over-all process, relative to the corresponding activation energies for gas phase reactions. All types of heterogeneous catalysis proceed via adsorbed intermediates, and there are two ways in which adsorption can lead to a lowering of activation energies.

First, the surface may act mainly to dissociate or weaken bonds in the adsorbate, so that the resultant ad-species can react with each other or with gas phase molecules. This occurs predominantly on metal surfaces, where, as we have seen, dissociative adsorption can take place without activation. Despite widespread study of such reactions, little information on the detailed mechanisms is available. At present, it seems that even the simplest reaction of this kind, i.e., H_2 - D_2 exchange, can occur by different mechanisms under different conditions. Thus, under conditions where the adsorbed atoms are mobile, the reaction occurs by surface mixing of hydrogen and deuterium atoms, but when the mobility of adsorbed hydrogen atoms is low, the ad-atoms may react with weakly adsorbed molecules (Gomer et al., 1957; Gomer, 1959). It is probable that some of the recent techniques developed for the study of chemisorption on metals (sec. 8.5) will help to elucidate mechanisms of this kind.

Secondly, the surface may act as a temporary donor or acceptor of electrons or holes. This generally occurs on semiconductors. In some cases, the catalyst may even function as a temporary donor or acceptor of protons. Some insulators, for example, are believed to contain acid centers which are capable of accepting protons in certain catalytic cracking reactions.

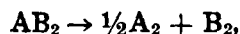
Catalytic reactions on all classes of materials can be thought of as involving the following sequence of individual steps:

- (1) Diffusion of the reactants to the surface
- (2) Adsorption of the reactants on the surface
- (3) Reaction
- (4) Desorption of the products from the surface
- (5) Diffusion of the products away from the surface.

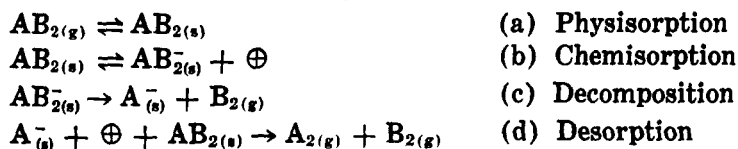
When any of these steps is slow compared to the others it may become rate controlling. Some industrially important processes, for example, become diffusion limited when finely-porous catalysts are used. Often, however, the overall rate and therefore the success of catalytic reactions depends on a number of factors quite apart from diffusion effects, which can often be eliminated. Thus, stringent requirements are usually placed on the energies of adsorption, reaction and desorption. If the activation energy for adsorption is high, reaction will be slow even if all other factors are favorable. Similarly, the activation energies for reaction and desorption must not be excessive or the reaction products either will not be formed or, if reaction occurs, the products may not be desorbed. Since these three steps—adsorption, reaction and desorption—are intimately related, very strong chemisorption is not desirable because reaction and desorption will then be too slow. On the other hand, weak adsorption will not influence the structures of ad-species sufficiently to lower the activation energies for reaction, although desorption of the products would be relatively rapid under such conditions.

We may well ask at this point why catalytic reactions occur at all, since these requirements seem very stringent. The answer lies partly in the fact that surfaces are rather heterogeneous with respect to the energies of adsorption. For metals, this heterogeneity arises partly from structural factors, i.e., the detailed atomic configuration at the surface; for semiconductors, as we have seen, adsorption influences the local electronic structure and hence all subsequent adsorption behavior. Thus, catalytic reactions may become self-regulating to some extent, and will proceed when the proper degree of adsorption has occurred.

This situation may be illustrated with reference to a simple decomposition reaction on a p-type semiconductor. The overall reaction is assumed to be



which may proceed according to the following sequence:



The subscripts (g) and (s) denote gas phase and adsorbed species, respectively, and the symbol \oplus denotes a positive hole in the space charge region. If steps (a) and (c) occur relatively rapidly, then steps (b) or (d) may become rate controlling. Considerations

similar to those discussed in section 8.1.4 indicate that an increase in the hole concentration, or a decrease in the Fermi level will accelerate step (d) and retard step (b). If the concentration of holes is sufficiently high, step (b) will therefore become rate limiting.

The overall reaction rate will be a maximum when the rates of reactions (b) and (d) are equal. This can be achieved by adjusting the Fermi level through the introduction of impurity ions. Thus, the overall rate of reaction on NiO may be lowered by the addition of Li_2O or raised by the addition of Ga_2O_3 or In_2O_3 . By considerations of this kind, Hauffe and Schlosser (1957) were able to interpret the rates of decomposition of N_2O on n- and p-type catalysts.

Two other phenomena of considerable practical importance, i.e., ageing and poisoning of catalysts, deserve mention. Ageing may have various origins. On semiconductors, it may be associated with field transport of ions in the space charge region, as already mentioned in connection with chemisorption. In some cases, particularly on metals, ageing may be due to preliminary reactions which alter the characteristics of the surfaces. Thus, it is even possible that initially clean metal surfaces are converted into semiconductors by the formation of superficial carbide or oxide films. Alternatively, ageing may consist merely in a roughening up of the surface. Thus, the catalytic combination of H_2 and O_2 on Cu or Ni seems to involve, at least initially, some surface oxidation; the subsequent reduction of this oxide layer results in a greatly altered or roughened surface.

Poisoning appears to be a complex and little understood phenomenon (Dowden, 1958). In some instances, such as the inhibition of hydrogen adsorption on metals by oxygen, poisoning may simply involve the competition between strongly and less strongly bound species. Even here it is possible, however, that adsorbed oxygen atoms prevent the unactivated dissociation of hydrogen molecules rather than the adsorption of hydrogen atoms. On semiconductors, poisoning may be associated with the formation of new surface states or the saturation of existing ones in a way unfavorable to the desired reactions. Since poisons seem to be effective in less than monolayer amounts, it is possible that they are preferentially adsorbed at active sites for reaction. Definitive information on the role of surface structure and structural imperfections in adsorption and catalytic reactions will be required before poisoning phenomena can be understood. Such information is lacking at the moment.

The ultimate aim of catalytic investigations is the prediction of the activity of catalysts and the development of new catalysts

from purely theoretical considerations. At present, we are far short of achieving these goals. Although theory provides a point of departure, most catalysts are still being prepared and evaluated by empirical methods. To achieve the long-range objectives, it first will be necessary to obtain a much better understanding of the mechanisms of currently known reactions and of the factors which affect both the mechanisms and kinetics of such reactions.

Information is needed not only on the rate constants and activation energies of the elementary steps in catalytic reactions, but also on the mechanistic nature of these individual steps. It also seems essential that the role of structural imperfections and of the surface structure, in general, be established. Moreover, it is not known to what extent catalytic reactions are controlled by surface diffusion. Systematic study of the relations between the electronic structure of metals and semiconductors and their catalytic behavior is also required. Finally, it will be appreciated that the study of catalysis cannot be divorced from the study of chemisorption alone and that all of the uncertainties in our present understanding of chemisorption apply to catalysis as well.

It is believed, however, that much can be accomplished with existing techniques. For example, tracer experiments can be used to elucidate the elementary steps of reactions and such experiments would be highly desirable, especially if carried out on single crystal surfaces of controlled orientation and structural perfection. Experiments of this type might profitably be extended and combined, where possible, with electrical conductivity and work function measurements. In all studies of the relation between catalyst structure and catalytic activity, it is vital that the surface structure of the catalyst be investigated by diffraction and other techniques both before and after reaction has occurred. It is possible, at present, to study the surface mobility of adsorbates on metallic substrates (Gomer and Hulm, 1957; Gomer, 1959R), but no work of a similar nature has yet been carried out on semiconductors. Field emission techniques could conceivably be used for this purpose, utilizing whiskers as emitters (Melmed and Gomer, 1961).

More systematic studies of the dependence of catalytic activity on the Fermi level, the electron or hole concentrations and the band edge energies in metals and semiconductors would be highly informative. It is possible to vary these factors by the addition of impurity ions in semiconductors (Engell, 1954R) and by alloying in metal systems (Schwab, 1953R). If these experiments are to have any real value, however, a number of important

criteria must be met. First, all other significant variables (i.e., surface area, structure, etc.) must be carefully controlled and held as nearly constant as possible. Second, the investigations should be designed in such a manner that the changes in rate constants and activation energies of the elementary steps can be determined rather than the change in the overall reaction rate alone. The rate constant for the overall reaction is usually a complex function of the individual rate constants, so that it becomes exceedingly difficult to arrive at an interpretation of mechanism from overall rate data.

There is also room for designing new types of experiments. Some of these are suggested in the section on modern experimental methods (sec. 8.5). Among additional possibilities, mention may be made of the evaporation of thin layers of catalysts onto substrates such as SiO_2 or Al_2O_3 . This would produce changes in the electron concentration of catalyst and substrate, since their Fermi levels differ initially. A double layer boundary would then be observed at the catalyst/substrate interface, the properties of which should influence the catalytic behavior profoundly.

In summary, it appears that future progress in this field will hinge largely on the adoption of new techniques and approaches based on modern concepts of solid state physics. The planning and execution of meaningful or significant experiments will therefore require close cooperation with specialists in the fields of semiconductors and metals.

8.3. Gas-Solid Reactions

Greatest progress in our understanding of the mechanism of gas-solid reactions has come from the study of the oxidation of metals and alloys and from investigations on related processes such as the tarnishing of metals by reaction with halogens, sulfur or hydrogen sulfide, etc. It is known that at high temperatures and under conditions where a dense, continuous and adherent oxide layer is formed on the surface of the metal, diffusion through the oxide layer is often rate determining (Hauffe, 1956R; Jost, 1952R; Wagner, 1951R). This is manifested by the appearance of a parabolic rate law. In certain circumstances, i.e., when diffusion through the oxide layer is relatively rapid, reactions at either the oxide/gas or metal/oxide interface may become rate controlling even when a dense, continuous oxide layer is formed. For example, in the oxidation of iron in $\text{CO-CO}_2\text{-O}_2$ atmospheres, the rate of chemisorption of oxygen from the gas phase onto the surface of the oxide appears to limit the rate of oxidation in certain

temperature and pressure ranges (Hauffe and Pfeiffer, 1952). Situations in which phase boundary reactions are rate controlling at high temperatures have not been observed very frequently, however (Hauffe, 1956R).

At low and intermediate temperatures, additional effects become important, in particular the presence of space-charge regions at the metal/oxide and oxide/gas interfaces (Cabrera and Mott, 1949R; Engell, et al., 1954; Hauffe, 1956R). These give rise to other mechanisms of oxidation and lead to nonparabolic rate laws. In addition, stresses are developed in the oxide layer as it thickens and if plastic flow at the reaction temperature is insufficient to accommodate or relieve these stresses, a system of cracks may develop in the oxide layer causing it to lose its protectiveness. Flaking or spalling may then be observed and the oxide layer may become physically detached from the substrate over a substantial area of the specimen. Mechanical breakdown of the oxide layer is frequently accompanied by a transition from a parabolic to a linear rate law and in many cases this transition usually occurs at some fairly definite thickness, the critical thickness for a given system depending primarily upon the oxidation temperature.

Transport or diffusion through a dense continuous oxide layer requires that there be chemical concentration or activity gradients in the oxide and this necessarily implies the existence of departures from ideal stoichiometry in the oxide itself. Even when a simple diffusion-limited mechanism is operative, the number of possible diffusion paths is quite large. Thus, departures from stoichiometry may be associated with the presence of various kinds of lattice defects in either the metal ion or the oxygen ion sublattices or both. In most oxides, sulfides, etc., the defects present in thermal equilibrium consist of vacant lattice sites (Schottky defects) or lattice ions displaced into interstitial positions (Frenkel defects). Depending upon the nature and concentration of lattice defects which are present, mass transport may involve either the diffusion of anions or cations and this may take place by either a vacancy or an interstitial mechanism (Jost, 1952R; Hauffe, 1956R); moreover, ionic transport must be accompanied by electron or hole conduction as to maintain electrical neutrality. Since the defect structure of the oxide will be influenced by the pressure of reacting gas and by the presence of impurity ions, the limiting diffusion rate and hence the rate of oxidation will depend sensitively on these various factors. Experiments with both pure metals (Wagner, 1951R; Hauffe, 1956R) and with dilute alloys (Gensch and Hauffe, 1950; Wagner and Zimens, 1947) have con-

firmed the essential correctness of this general picture of high temperature oxidation.

The oxidation behavior of more concentrated alloys at high temperatures is also frequently governed by a parabolic rate law. However, the mechanism of oxidation is even more complicated since the following additional factors must be taken into account (Wagner, 1959).

(1) In general, the two constituents of a binary alloy have different affinities for oxygen and their ions have different mobilities in the oxide layer. Consequently, a more or less pronounced enrichment of one constituent may occur in certain regions of the oxide layer.

(2) If preferential or selective oxidation of one constituent occurs, and if the mobility of this constituent in the oxide layer is sufficiently high, the rate of diffusion of this component in the alloy itself may become rate controlling.

(3) When both components are oxidized and the (different) mobilities of the constituent ions in the oxide layer are both less than the corresponding diffusion rates in the alloy, two oxide layers of differing thickness (and structure) are frequently observed.

(4) If the reacting gas is soluble in the alloy (e.g., oxygen in copper- or silver-base alloys), and if it diffuses at a faster rate in the alloy than the most reactive alloy constituent, oxide formation in the interior of the alloy (internal oxidation) will accompany the growth of the external oxide layer (Meijering and Druyvesteyn, 1947a, 1947b; Rhines, 1951R; Wagner, 1959).

At intermediate and low temperatures, i.e., in the so-called "thin film" region, a more complicated mechanism replaces the ambipolar diffusion mechanism responsible for the parabolic rate law which governs the growth of "thick films" or scales at high temperatures. This change in mechanism is brought about by the presence of steep gradients in the chemical and electric potentials at the phase boundaries; depending upon the conditions, either a cubic, a parabolic or some form of logarithmic rate law may be obtained. The parabolic rate law observed at low temperatures when the oxide film is relatively thin may be attributed to the field transport of ions through the film. Field transport results from the formation of a space charge region in the oxide due to chemisorption of the reacting gas (Engell and Hauffe, 1952), a phenomenon which has already been encountered in connection with chemisorption on semiconductors. At low temperatures ion transport through the film occurs only with the aid of the space charge

field. Quite apart from the experimental difficulties of investigating field transport phenomena of this kind, attempts to formulate exact rate laws from the general transport equations lead to as yet unsolved integro-differential equations.

The interpretation of the cubic and logarithmic rate law currently presents the greatest difficulties (Cabrera and Mott, 1949R; Engell et al., 1954; Hauffe, 1956R, 1959R), although a mechanism which accounts for the latter has recently been proposed (Hauffe and Schottky, 1958R).

Although the theory outlined here is probably valid in its major assumptions for dense, continuous films, it is seriously incomplete. Recent work indicates, for example, that continuous films may be the exception rather than the rule, and that discrete oxide nuclei are often formed in the early stages of reaction (Benard, 1960; Gulbransen and Copan, 1959). Furthermore, whiskers and thin sheet-like protrusions (approximately 100Å thickness) are often seen growing outward from the surface in the "thick film" region (Pfefferkorn, 1953; Gulbransen and Copan, 1959). All this suggests that structural considerations play an exceedingly important role. At present, however, it is not known with certainty whether oxide whiskers are nucleated at structural imperfections in the metal or in the oxide (Young, 1960) or at regions of mismatch in the interface, and whether they grow from the tip or the base. A more complete theory of oxidation will have to consider nucleation problems in terms of atomic structure and imperfections and also the factors which determine whether film formation is continuous or not. Clearly a great deal of experimental information is still needed before this can be attempted.

Reduction and roasting processes also belong in the domain of gas-solid reactions. Reactions of this type are not nearly as well understood as are oxidation processes in general. All such reactions are initiated by chemisorption of the corresponding gas. Hence it is particularly desirable to study the kinetics of the uptake of CO, H₂ and O₂ by oxides whose defect structure is known. Work of this kind exists, but must be extended in the light of semiconductor theory. Since chemisorption is often connected with a certain amount of surface reaction, the study of desorption is interesting for two reasons: First, one can find out how much of the initial reactant is still chemisorbed. Second, the desorption kinetics of the gaseous end products are important for optimizing the overall reaction.

In addition to kinetic and nucleation studies, calorimetric measurements of adsorption and reaction heats are important if indi-

vidual quantities can be determined by a proper choice of experimental parameters.

The progress of a reduction can be followed not only by volumetric gas analysis but also by resistivity and X-ray measurements. Without doubt the details of nucleation and nucleus growth are least understood in this field.

8.4. Impact Phenomena

Collisions of molecules, atoms, and ions with surfaces are of considerable scientific and technological interest. This section deals briefly with sticking and thermal accommodation coefficients, and with dissociation, recombination and sputtering. While these topics are pertinent to adsorption, it is convenient to discuss them collectively at this point. An account of work through 1950 is given by Massey and Burhop (1952R), and will not be reviewed here.

With the possible exception of sputtering at high energies, impact phenomena depend sensitively on the potential seen by the impinging particle. This has been mentioned briefly in connection with adsorption. It should be emphasized that the one-dimensional approximations which are useful in treating certain aspects of adsorption are inadequate for handling most collision problems. Unfortunately, relatively little experimental or theoretical information on the details of the three-dimensional surface potential is presently available. In addition, the mathematical difficulties involved in treating collisions are so formidable that the models employed up to the present time have been grossly over-simplified and are inadequate in many important respects. Of necessity, therefore, the following discussion will be rather qualitative.

8.4.1. Thermal Accommodation

When the mean free path of molecules in a gas exceeds the dimensions of the containing vessel, gaseous heat transfer from a hot to a cold surface proceeds by direct molecular transport and depends on the extent to which impinging molecules are "heated up" on the hot surface. The accommodation coefficient β is a measure of the efficiency of thermal transfer at the hot surface. It is defined as the fraction of energy gained by the departing molecules relative to that corresponding to complete equilibration with the hot surface.

Numerous experimental measurements of β have been made under a variety of conditions, most of which yield values of β of the order of 0.1 or less on many surfaces. It is more illuminating,

however, to convert the energy carried away into an evaporation temperature, T_e , by assigning a reasonable heat capacity to the molecules. This has been done for a variety of gases on graphite (Meyer and Gomer, 1958). It is then found that T_e is constant, independent of the temperature of the hot surface, T_s , and of that of the cold walls, T_w , as long as $T_w < T_e < T_s$. This suggests the following mechanism of thermal transfer. An incoming molecule transfers some of its kinetic energy rapidly (in a time less than a vibrational period) to the surface or to vibrational modes other than those leading to desorption or re-evaporation. Desorption occurs only when all the vibrational modes involving the molecule, including the desorption mode, have been excited to an equivalent temperature, T_e . At present, it seems difficult to account for the constancy of kT_e on any other basis.

This mechanism indicates two things. First, energy loss to the solid is extremely rapid. Second, the lifetime of an impinging molecule on the surface exceeds one vibrational period and is, in fact, sufficiently long to permit a quasi-equipartition of energy among its internal as well as translational modes. The first conclusion is in complete agreement with a semi-classical calculation recently performed by Cabrera (1959), based on a linear chain of harmonically-coupled atoms, while the second is a plausible consequence of the three-dimensional nature of the potential and the anharmonic coupling of the resultant modes.

Unfortunately, all existing quantum mechanical theories ignore these facts completely and are based on a transfer mechanism which is confined to one vibrational period (Lennard-Jones, 1937).

As pointed out by Cabrera (1959), the original energy transfer to the surface may occur by a phonon cascade mechanism rather than by a one-phonon transition. The indications are that this may also be true for the reverse process. Thus, many phonons may diffuse into the surface-adparticle complex before desorption takes place. The presently available theories are therefore applicable only when the energy transfer occurs in a single collision, such as seems to be the case when T_w is greater than T_e but less than T_s . Unfortunately, most experimental work carried out to date has not taken these considerations into account, so that it is difficult to interpret any given set of data.

There is a great deal of evidence which indicates that β varies with surface coverage of chemisorbed gases (Roberts, 1935, 1937). Most probably, β is also structure-sensitive and depends upon the crystallographic orientation as well, although little experimental information exists on this point. The dependence of β

on surface coverage has never been explained satisfactorily; it might possibly be connected with a reduction in T_s or with an inhibition of energy loss by the impinging molecule.

It is clearly desirable to obtain more experimental information on the mechanism of thermal accommodation, using the approach outlined by Meyer and Gomer (1958). Existing techniques make it feasible to carry out such experiments with a number of simple substrates and gases. It would also be interesting to perform experiments with single crystal ribbon substrates in order to assess the effect of surface orientation. Furthermore, it is also clear that presently available theoretical treatments are quite inadequate and that methods must be developed for handling more realistic models.

8.4.2. Condensation

When $T_s \sim T_e$, condensation can occur. Under these conditions, the problem of interest is to evaluate the sticking coefficient, α , which may be defined as the fraction of impinging particles that remain on the surface for a time comparable to the mean life, τ , of an adsorbed particle. This definition does not, of course, exclude the case where $T_s > T_e$, for which τ is very short. In practice, however, α is determined only when τ is effectively infinite.

In the simplest situations, where there is only one ad-state and dissociation does not occur, α and β are closely related, since both reflect the rate of energy transfer from the impinging particle to the solid. As previously noted, theoretical calculations are rather unreliable, although even the linear chain model used by Cabrera (1959) indicates that α is close to unity. Again, the experimental picture is very incomplete. The experiments of Meyer and Gomer (1958) on thermal accommodation certainly indicate that a rapid energy transfer occurs. There is also evidence, from the kinetics of Hg whisker growth, that the sticking coefficient of Hg atoms on an Hg surface at -78°C is close to unity (Gomer, 1958). It would be highly desirable to measure sticking coefficients for monatomic, physically adsorbed gases on metals. For instance, it is quite within present capabilities to perform flash filament experiments (sec. 8.5.6) with inert gases on metals such as tungsten.

In recent years, information on the sticking coefficient of chemisorbed gases on metals has been obtained by the flash filament technique. It is found (Kisliuk, 1957) that α lies between 0.1 and 0.5 (there is still some experimental discrepancy here) at low coverage, and decreases both with increasing coverage and increasing temperature. These facts clearly indicate a fairly complex

mechanism, the general features of which can probably be described as follows: impinging molecules are physically adsorbed and wander freely over the surface. Since the mean life for the physisorbed state is very short at the temperatures of the experiment, re-evaporation occurs unless a wandering molecule can become chemisorbed. Different expressions for the variation of the overall sticking coefficient with θ and T can be formulated, depending on the details of this mechanism. For instance, it may be assumed that physically adsorbed molecules wander randomly over the surface for a time determined by a single desorption energy and frequency, regardless of whether any chemisorption site which the molecule encounters is filled or empty. In this case, α is given by (Gomer, unpublished)

$$\alpha = \alpha_p [1 - (1 - \alpha_c)^{(1-\theta) \exp[(E_{des} - E_{diff})/kT]}] \quad (8.3)$$

where α_p is the sticking coefficient for physical adsorption, α_c the chance that a physically adsorbed molecule which encounters an empty chemisorption site will become chemisorbed, and E_{des} and E_{diff} are the activation energies of evaporation and surface diffusion respectively. If it is assumed, on the other hand, that evaporation of a physically adsorbed molecule occurs when it encounters a filled chemisorption site, the overall sticking coefficient is given by Gomer (unpublished):

$$\alpha = \alpha_p (1 - \theta) [1 - (1 - \alpha_c)^{1/\theta}] \text{ for } 1/\theta \leq \exp[(E_{des} - E_{diff})/kT] \quad (8.4)$$

Eqs. (8.3) and (8.4) indicate that $\alpha_c < \sim 10^{-2}$ if $\alpha_p = 1$.

At present, little information is available on the variation of α or its components, α_c and α_p , with surface orientation or structure. It would be of interest to carry out flash filament experiments on a number of single crystal ribbons of different surface orientation. A sound theoretical expression for α is also lacking. As pointed out in the section on chemisorption, α_c usually does not involve an activation energy, a fact which can be rationalized more or less successfully. Beyond that little can be said about this quantity.

8.4.3. Dissociation and Recombination

It was discovered by Langmuir (1912) that hot tungsten filaments are capable of dissociating hydrogen molecules into gaseous atoms even at very low pressures. Conversely, the recombination of atoms to yield molecular species on suitable catalysts has also been known for a long time. Since these phenomena are closely related to sticking, they will be considered here.

The conversion of molecular to atomic species most probably involves dissociative chemisorption followed by atomic evaporation. As indicated in the discussion of chemisorption, this will occur only when the filament temperature is so high that the rates of these processes are competitive with molecular desorption. Very little is known about the absolute efficiency of atom formation. It would appear, however, that the efficiency should be calculable from the initial sticking coefficient of the molecular species and the relative rates of atomic and molecular desorption. Consequently, on raising the temperature, it might be expected that maximum efficiency would be obtained at a temperature which represents the best compromise between decreasing sticking probability and increasing rate of atomic desorption. It is interesting, in this connection, that hot graphite filaments will dissociate CH_4 at pressures where a sheath of hot gas forms, but not under conditions where only surface impacts occur (Meyer and Gomer, 1958).

The inverse process, recombination, has been studied most extensively with hydrogen, although some work on the recombination of oxygen atoms has also been reported. It appears that metals on which hydrogen is strongly chemisorbed are particularly efficient in catalyzing recombination. Existing experimental information is scanty and the data are exceedingly crude. Frequently, the experiments have sought only to compare the efficiencies of recombination of various metals and, for this purpose, only the relative temperature rise of the substrate (resulting from the exothermic nature of the recombination) may have been measured. A few measurements of absolute efficiencies do exist, however (Wood and Wise, 1958); the values obtained vary from 0.25 for Pt to 10^{-3} for Al (undoubtedly oxide covered) and 8×10^{-4} for Pyrex glass.

Detailed kinetic data are lacking at present. It is probable, however, that the formation of molecular species involves either the surface recombination and molecular desorption of adsorbed atoms, or the reaction of an ad-atom with an atom impinging on the surface from the gas phase. The situation is somewhat reminiscent of the H_2 - D_2 exchange reaction. The first mechanism may be operative when ad-atoms are mobile and when the surface temperature is sufficiently high to permit desorption of the molecular species. Again, a definite maximum in the recombination rate would be expected at a temperature where: a) the mobility of ad-atoms is high; b) molecular desorption is rapid; c) the rate of atomic desorption is still low; and d) the steady-state surface coverage is sufficiently low to make the initial sticking coefficient high. It is seen that a great deal of work needs to be done here. It

would be desirable to measure absolute efficiencies as a function of atom concentration, temperature and crystallographic orientation. Until this is done there is little point in further speculation about the mechanism. Attempts to correlate recombination efficiency with the electronic structure of the substrate have been made (Pickup and Trapnell, 1956) in the past. While it is undoubtedly true that the ability to chemisorb hydrogen is important here, the complexity of the overall reaction, the uncertainties in the experimental results, the possible presence of unscavenged oxide films, the pitfalls in many H_2 adsorption data, and so on, make these attempts somewhat premature.

Nevertheless, dissociation and recombination are scientifically and technologically interesting. One or both may play a fairly important role during some phases of missile re-entry, and in problems concerning satellites and space vehicles. It is hoped that more work will be forthcoming (Ehrlich, 1959).

8.4.4. Sputtering

Cathode sputtering consists of the removal of target material by ion bombardment. Discussion here is limited to physical effects and does not consider chemical sputtering, which can occur under similar experimental conditions but consists of the formation and subsequent decomposition of volatile compounds by surface reactions. Sputtering is technologically important because it is a method of cleaning surfaces, because it may contribute to meteor and space vehicle ablation, and because its occurrence in gas discharge tubes often leads to undesirable side effects. It is also of considerable scientific interest, as discussed below. Despite the very large amount of experimental work performed since the middle of the 19th century, it is only recently that a working hypothesis of its mechanism has been proposed. This is due to the fact that fairly stringent experimental conditions must be met before results become meaningful for even such gross features as the sputtering yield. It is necessary to use mono-energetic ions, preferably of known impact direction, well characterized surfaces, and to prevent competing attrition processes such as oxide formation and removal. An excellent summary of work through 1955, indicating these pitfalls and means of overcoming them, is given by Wehner (1955R); see also Part XI, section 4.

The following experimental facts seem established. Sputtering consists mainly of the removal of single, energetic atoms (10-20 eV) whose velocities are directed away from the incident beam. Sputtering leads to profound surface rearrangement, both on the

atomic and on a macroscopic scale. Crystallographically oriented etch structures frequently occur.

There is a definite energy threshold for sputtering in a given system. Above this threshold, the sputtering yields vary at first parabolically and then linearly with ion energy. The threshold and yields depend both on the substrate parameters and the mass of the impinging ion and have been combined into an empirical formula by Wehner (1955R).

These facts preclude some of the older theories, such as the hot-spot hypothesis, which assume that local melting and evaporation occur at the locus of ion impact. It seems much more probable, at least at low energies, that direct binary collisions, resulting in momentum and energy transfer, are involved. The details are by no means established yet, although two recent theories (Henschke, 1957; Langberg, 1958) seem to account reasonably well for the few experimental results presently available. Neither theory is sufficiently detailed to predict the formation of etch structures, although it is quite clear that the ejection of lattice surface atoms must depend on structure and orientation.

Further advances in the theory of sputtering will hinge on the availability of more experimental information. In addition to overall yield, data on the angular and energy distributions of ejected atoms and ions are needed. Experiments with oriented single crystal faces, particularly to determine the effect of sputtering on their structure, would be useful. Recent work by Müller (unpublished) with the field ion microscope indicates one approach. More experiments on the effect of impact angle (by collimated beam techniques) are needed. Information on the fate of the impinging ion is lacking and it would be very desirable to establish, by means of tracer experiments, what fraction recoils and what fraction burrows into the lattice. More attention should be given to high energy sputtering and to the ejection of ions, which has been observed (Bradley, 1958R).

Knowledge of the interaction potential between the lattice atoms and the impinging ion is needed; for example, it is not always known whether and where the impinging ion is neutralized. This is a similar problem to that encountered in secondary electron emission by ions and metastable atoms, and has been treated theoretically for only a few simple cases (Cobas and Lamb, 1944).

8.5. Modern Experimental Techniques

Since every conclusion in the field of gas-solid interaction hinges sensitively on proper experimental conditions, it is worthwhile to

discuss briefly some of the newer techniques which have been developed and applied, in recent years, to the study of surface phenomena (also, see Part XII on "Instrumentation and Techniques").

8.5.1. Ultra-High Vacuum Techniques

Attainment of pressures of 10^{-11} mm or better in glass or metal systems can be achieved readily if the entire system can be baked out under vacuum at about 430°C for a few hours. Since heats of desorption increase with decreasing coverage, the removal of the more labile ad- and absorbate fractions at high temperatures insures vastly reduced desorption rates at low temperature. Recent improvements in this technique (which was employed successfully 30 years ago) include the insertion of copper foil getters² in the vacuum path, ion pumping, and bakeable all metal valves which facilitate clean gas handling (Alpert 1953a, 1953b; Alpert and Buritz, 1954; Pollard, 1959R). Thin-wire-collector ion gauges, radio frequency mass spectrometers of small size (Alpert 1953a, 1953b; Alpert and Buritz, 1954) and monitor field emission tubes (Good and Müller, 1956R; Gomer, 1961R) have it possible to estimate pressures below 10^{-10} mm and to analyze the composition of the gas residues with relative ease.

An entirely different approach, and in some ways a simpler one, takes advantage of the low vapor pressures of all gases except He at 4.2°K . Liquid He or H_2 cooling (Gomer et al., 1957; Wortman et al., 1957; Gomer and Hulm, 1957; Gomer, 1961R) has been used extensively in field emission work and leads to pressures of $< 10^{-15}$ mm.

8.5.2. Cleaning of Surfaces

Although high vacuum is necessary for maintenance, it has little bearing on the initial attainment of clean surfaces. Achievement of high purity surfaces usually involves thermal desorption, ion sputtering (Farnsworth and Woodcock, 1957R; Wehner, 1955R), or the creation of a new surface by evaporation or by vacuum cleavage. The first of these methods works for high melting metals like W, Mo, etc., but becomes difficult for Ni and almost impossible for metals like Cu and for many oxides and

² The action of copper foil getters depends on the fact that heating to 430°C results in pseudo-distillation of Cu (by labile compound formation and subsequent decomposition) and the consequent formation of enormous real surface areas capable of adsorbing vast amounts of gas.

sulfides. Partly, this is because the volatility of surface phases is either too low or too high at the maximum feasible temperatures, and partly because the diffusion of impurities from the interior leads to the formation of epitaxed nonvolatile surface phases. This can be seen very clearly with the field emission microscope (Good and Müller, 1956R; Gomer, 1961R). Sputtering suffers from the disadvantage that the resulting surface is atomically roughened by this rather violent treatment so that thermal annealing, with the concomitant danger of diffusion from the interior, occurs. Evaporated films are usually free from these drawbacks but have the disadvantages of varying and often poorly-definable crystallographic structure. Vacuum cleavage has been used relatively infrequently but can produce reasonably clean and moderately uniform surfaces; however, this technique cannot be generally applied to all classes of materials.

8.5.3. Field Emission Microscopy

The field emission microscope, invented by E. W. Müller, consists of a small monocrystalline metal tip of 10^{-6} to 10^{-5} cm radius surrounded by a fluorescent screen anode. In fields of the order of 3 to 5×10^7 volts/cm, emission of electrons occurs from the tip. The radially divergent beam produces on the screen a highly magnified electron emission map of the tip surface, showing the crystal symmetry. Since adsorption produces changes in work function, even small fractions of monolayers show up by enhanced or decreased emission. The emitters used consist of relatively perfect single crystals of known orientation and surface condition and make ideal substrates for many adsorption studies. The technique has been employed to obtain (Becker, 1955R) ; Gomer, 1961R; Good and Müller, 1956R) : work function measurements as function of orientation; estimates of relative coverages from work function changes; kinetics of adsorption; rates and activation energies of desorption; quantitative measurements of surface diffusion rates and activation energies.

The disadvantages of the method are the necessity of employing high fields (which may only be pulsed) for observation (but not occurrence) of processes, and the difficulty of preparing emitters from materials with low melting points. The latter difficulty seems to have been at least partially overcome by the use of whiskers, grown, *in situ*, from the vapor under high vacuum conditions. Thus, Melmed and Gomer (1961) have prepared clean Cu, Fe, Al, Ge, Au, Ag, Hg, etc. field emitters by this method. Similar investigations with oxides and sulfides are difficult and as yet lacking.

8.5.4. Field Ion Microscopy

The field ion microscope is a modification of the field emission microscope and is capable of much higher resolution ($\sim 3\text{\AA}$); thus, it is suitable for structural investigations (Gomer, 1955R; Good and Müller, 1956R; Müller, 1960). It consists of a field emission tube containing He or other gas under a few microns pressure; the metal tip is the positive electrode. At fields of the order of $2\text{--}4\text{v/\AA}$, gas molecules or atoms finding themselves near the tip can lose an electron by field emission from the molecule into the tip. This process is called field ionization. The positive ions so formed follow trajectories identical to those of electrons in field emission and produce a magnified image of the tip on the screen. At fields just high enough to permit ionization, the small variations in field arising from the discrete atomic structure of the tip affect the ionization rate critically, so that the image reproduces the potential and hence the physical structure of the lattice quite closely. The resolution is determined by the thermal velocity of the ions, when these are formed on the rebound from atoms which have been thermally accommodated on the cold tip. Very high resolutions therefore can be obtained by cooling the tip.

The method is useful only for materials able to withstand the high electrostatic stresses resulting from the high fields. It is possible, at the expense of resolution, to use gases with relatively low ionization potentials and thus to decrease the fields necessary for image formation.

Field ion microscopy is capable of resolving lattice planes and, sometimes, individual lattice atoms and is useful, therefore, in the study of imperfections, overgrowths, etc.

8.5.5. Field Desorption

Desorption from field emitters under the influence of very high positive fields has been studied by Müller (1956) and by Inghram and Gomer (1955) on metals. The phenomenon consists of direct evaporation of ions over a Schottky saddle for ad-particles whose ground state is effectively ionic (Müller, 1956; Gomer, 1959). For covalent binding, field desorption occurs by vibrational excitation to the point where the ground state curve is intersected by the ionic curve (which is bent downward in the presence of an electric field), followed by field ionization (Gomer, 1961R).

Analysis of the desorption rates, as a function of field and temperature, makes it possible not only to determine the original ground state but also to probe the shape of the corresponding

potential curve. In addition, mass spectrometric analysis of the desorption products (Inghram and Gomer, 1955) yields direct information on the ad-species, and, thus, provides a means for determining the products of dissociative chemisorption, even in complex systems. The experimental difficulties of mass spectrometric analysis in conjunction with this technique are considerable, but not insuperable.

8.5.6. Flash Filament Techniques

This technique, suggested by Apker (1948) and subsequently developed by Becker (1955R), Ehrlich (1961R) and others, consists of permitting adsorption on a clean, preferably monocrystalline, metal ribbon and then raising it to the desorption temperature(s). The pressure rise, or the actual desorption products, are then determined with an ion gauge or omegatron. In this way, rates of ad- and desorption can be investigated. The method has been most useful for determining kinetics of adsorption and sticking coefficients (Kisliuk, 1957). It has also been used by Ehrlich (1956) to investigate desorption temperatures and hence, qualitatively, desorption energies. In this application it lacks the resolution of field emission. On the other hand, the method is eminently suited for determining the absolute amounts of ad- or desorbed species and thus complements the field emission technique.

8.5.7. Contact Potential Determinations

Many techniques have been used for this purpose, but all have certain advantages and disadvantages. The most useful seems to be the vibrating condenser method developed by Mignolet (1955), since it permits correlation of the amounts adsorbed with work function changes and does not involve high temperatures or electron bombardment. It is applicable to both metals and semiconductors. Its chief disadvantage lies in the fact that the evaporated films used as adsorbents are of necessity porous and polycrystalline, so that the correlation of coverage with the amount adsorbed depends on diffusion into the interior of the films, which leads to uncertainties. Nevertheless, maximum contact potentials can be obtained very accurately with this technique.

The use of field emission for contact potential determinations has previously been mentioned. The results obtained agree with other methods and, moreover, there is the advantage of direct observation of the state of the specimen during the measurement, so that vitiating side effects become obvious. It is further

possible to obtain, on the same specimen and under a given set of conditions, work functions from individual crystal directions by measuring the emitted current from the separate faces. It is necessary, however, to carry out this refinement by electrical (e.g., Faraday cages with small holes) rather than photographic means to obtain high accuracy. The disadvantage of the field emission technique is that it is impossible to obtain the number of adsorbed molecules directly. Combination with the flash filament technique, either directly or by a separate determination of sticking coefficients, is therefore necessary.

Photoelectric work function determinations are also very useful in adsorption work. As pointed out in a previous section, the detailed theory of the photo-effect is still quite poorly understood. Nevertheless, experiments in which a metal ribbon is used both as a flash filament and as a photoelectric cathode have been carried out (Eisinger, 1958) and seem very promising.

8.5.8. Low-Energy Electron Diffraction

Recent work, largely by Farnsworth and his collaborators (Farnsworth and Woodcock, 1957R), promises to throw light on the actual atomic or crystallographic configuration of adsorbed monolayers. The principal disadvantages of low energy electron diffraction techniques are the difficulties in complete cleaning of large surfaces by sputtering or heating, and the buildup of epitaxial overgrowths, often covering only a small part of the surface, which generally result from incomplete heat treatment of a metal surface.

8.5.9. Infra-Red Spectroscopy

Eischens and his collaborators (Eischens and Pliskin, 1957R) have obtained infra-red spectra of chemisorbed CO and other molecules on powder surfaces and have deduced information on the binding and adsorbate-substrate geometry from these results. This method has promising potentialities although much work will be required before the conditions of cleanliness and surface specification necessary for detailed interpretation will be met.

8.5.10. Electron and Nuclear Resonance Methods

The application of these methods to surface problems is in its infancy (Pastor et al., 1957R); the difficulties inherent in infra-red spectroscopy are equally pertinent. Nevertheless these techniques represent potentially powerful tools for future studies of surfaces; see also section 7.2.5.

8.6. References

8.6.1. Reviews and General Articles

- Baker, M. McD., and Jenkins, G. I., 1955, in *Advances in Catalysis* (New York, Academic Press), VII, p. 1.
- Becker, J. A., 1955, in *Advances in Catalysis* (New York, Academic Press), VII, p. 135.
- Birchenall, C. E., 1959, in *Kinetics of High Temperature Processes*, W. D. Kingery, editor (Technology Press, MIT and John Wiley and Sons, New York), pp. 277-281.
- Bradley, R. C., 1958, AFSOR-TN-58-642.
- Cabrera, N., and Mott, N. F., 1949, Rep. Progr. Physics, 12, 163.
- Ehrlich, G., 1961, J. Chem. Phys., 32, 4.
- Eischens, R. P., and Pliskin, W. A., 1957, in *Advances in Catalysis* (New York, Academic Press), IX, p. 123.
- Engell, H. J., 1954, in *Halbleiterprobleme* (Braunschweig), Bd. 1, p. 249.
- Farnsworth, H. E., and Woodcock, R. F., 1957, in *Advances in Catalysis* (New York, Academic Press), IX, p. 123.
- Gomer, R., 1955, in *Advances in Catalysis* (New York, Academic Press), VII, p. 93.
- , 1961, Field Emission and Field Ionization (Cambridge, Mass., Harvard University Press).
- Good, R. H., and Müller, E. W., 1956, in *Handbuch der Physik* (Berlin, Springer-Verlag), XXI, p. 176.
- Hauffe, K., 1955, in *Reaktionen in und an Festen Stoffen* (Berlin, Springer-Verlag), p. 645.
- , 1956, *Oxydation von Metallen und Metallegierungen* (Berlin, Springer-Verlag), pp. 72-85 and pp. 214-227.
- , in *Kinetics of High Temperature Processes*, W. D. Kingery, editor (Technology Press, MIT and John Wiley and Sons, New York), pp. 282-293.
- , and Schottky, W., 1958, in *Halbleiterprobleme* (Braunschweig).
- Hill, T. L., 1953, in *Structure and Properties of Solid Surfaces*, R. Gomer and C. S. Smith, editors (Chicago, University of Chicago Press).
- Jost, W., 1952, *Diffusion in Solids, Liquids and Gases* (New York, Academic Press).
- Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford, Clarendon Press).
- Pastor, R. C., Weil, J. A., Brown, T. H., and Turkevitch, Jr., 1957, in *Advances in Catalysis* (New York, Academic Press), IX, p. 107.
- Pollard, J., 1959, Repts. on Prog. in Physics, Vol. XXII, pp. 33-74.
- Rhines, F. N., 1951, in *Atom Movements* (Cleveland, Ohio, American Society for Metals), pp. 174X191.
- Schwab, G. M., 1953, in *Structure and Properties of Solid Surfaces*, R. Gomer and C. S. Smith, editors (Chicago, University of Chicago Press), p. 474.
- Selwood, P. W., 1957, in *Advances in Catalysis* (New York, Academic Press), IX, p. 93.
- Suhrmann, R., 1955, in *Advances in Catalysis* (New York, Academic Press), VII, 303.
- Trapnell, B. M. W., 1954, Chemisorption (London, Butterworth's Publications Ltd.).

- Wagner, C., 1951, in *Atom Movements* (Cleveland, Ohio, American Society for Metals), pp. 153-173.
- Wehner, G. K., 1955, *Adv. in Electronics and Electron Physics*, 7, 239.
- Wheeler, A., 1953, in *Structure and Properties of Solid Surfaces* (Chicago, University of Chicago Press).

8.6.2. Original Articles

- Alpert, D., 1953, *J. Appl. Phys.*, 24, 860.
- , 1953, *Rev. Sci. Inst.*, 24, 1004.
- , and Buritz, R., 1954, *J. Appl. Phys.*, 25, 202.
- Apker, L., 1948, *Ind. and Eng. Chem.*, 40, 846.
- Benard, J., 1960, *Acta Met.*, 8, 272.
- Boudart, M., 1952, *J. Am. Chem. Soc.*, 74, 3556.
- Brunauer, S., Emmett, P. H., and Teller, E., 1938, *J. Am. Chem. Soc.*, 60, 309.
- Cabrera, N., 1959, *Disc. Faraday Soc.*, 28, 16.
- Cobas, A., and Lamb, W. E., 1944, *Phys. Rev.*, 66, 327.
- Dowden, D. A., 1958, *Bull. Soc. Chim. Belg.*, 67, 439.
- Ehrlich, G., 1956, *J. Phys. Chem.*, 60, 1388.
- , 1959, *J. Chem. Phys.*, 31, 1111.
- , Hickmott, T. W., and Hudda, F. G., 1958a, *J. Chem. Phys.*, 28, 977;
- 1958b, *ibid.*, 28, 506.
- Eisinger, J., 1958, *J. Chem. Phys.*, 29, 1154.
- Ely, D. D., 1950, *Disc. Faraday Soc.*, VIII, 34.
- Engell, H. J., and Hauffe, K., 1952, *Metall.*, 6, 285.
- , Hauffe, K., and Ilschner, B., 1954, *Z. Elektrochem.*, 58, 478.
- Garner, E. W., Gray, T. J., and Stone, F. S., 1949, *Proc. Roy. Soc.*, A197, 294.
- Gensch, C., and Hauffe, K., 1950, *Z. phys. Chem.*, 196, 427.
- Gomer, R., 1953, *J. Chem. Phys.*, 21, 1869; 1958a, *ibid.*, 28, 458; 1958b, *ibid.*, 29, 441; 1959, *ibid.*, 31, 341.
- , 1959, *Disc. Faraday Soc.*, 28, 23.
- , and Hulm, J. K., 1957, *J. Chem. Phys.*, 27, 1363.
- , Wortman, R., and Lundy, R., 1957, *J. Chem. Phys.*, 26, 1147.
- Gossel, H., and Hauffe, K., unpublished.
- Gray, T. J., and Savage, S. P., 1959, *Disc. Faraday Soc.*, 28, 159.
- Gulbransen, E. A., and Copan, T. P., *Disc. Faraday Soc.*, 28, 229.
- Halsey, G., 1948, *J. Chem. Phys.*, 16, 931.
- Hauffe, K., 1955, *Z. angew. Chem.*, 67, 189.
- , and Pfeiffer, H., 1952, *Z. Elektrochem.*, 56, 390.
- , and Schlosser, E. G., 1957, *Z. Elektrochem.*, 61, 711.
- Henschke, E., 1957, *Phys. Rev.*, 106, 737.
- Hickmott, T. W., and Ehrlich, G., 1958, *J. Phys. Chem. Solids*, 5, 47.
- Huang, K., and Wyllie, G., 1950, *Disc. Faraday Soc.*, VIII, 18.
- Inghram, M. G., and Gomer, R., 1955, *Zeitschr. f. Naturf.*, 10a, 863.
- Kisliuk, P., 1957, *J. Phys. Chem. Solids*, 8, 95.
- Kobayashi, H., and Wagner, C., 1954, *J. Chem. Phys.*, 26, 1609.
- Langberg, E., 1958, *Phys. Rev.*, 111, 91.
- Langmuir, I., 1912, *J. Am. Chem. Soc.*, 34, 1310.
- Lennard-Jones, J. E., 1937, *Proc. Roy. Soc.*, A163, 127.
- Medved, D. B., 1958, *J. Chem. Phys.*, 28, 870.
- Meijering, J. L., and Druyvesteyn, M. J., 1947a, *Philips Research Reports*, 2, 81; 1947b, *ibid.*, 2, 260.

- Melmed, A. J., and Gomer, R., 1961, *J. Chem. Phys.*, **34**, 1082.
Meyer, L., and Gomer, R., 1958, *J. Chem. Phys.*, **28**, 617.
Mignolet, C. P., 1955, *J. Chem. Phys.*, **23**, 753.
Müller, E. W., 1956, *Phys. Rev.*, **102**, 618.
———, 1958, *Acta Met.*, **6**, 620.
———, 1960, *Adv. Electronics and Electron Physics*, **13**, 88.
Pfefferkorn, G., 1953, *Naturwiss.*, **40**, 551.
Pickup, K. G., and Trapnell, B. M. W., 1956, *J. Chem. Phys.*, **25**, 182.
Rhines, F. N., 1940, *Trans. AIME*, **137**, 246.
———, 1947, *Corrosion*, **4**, 15.
Roberts, J. I., 1935, *Proc. Roy. Soc.*, **A152**, 445; 1937, *ibid.*, **A161**, 141.
Sandomirski, V. B., 1957, *Bull. Acad. Sci., USSR*, **21**, 211.
Schlosser, E. G., and Hauffe, K., unpublished.
Stevenson, D. P., 1955, *J. Chem. Phys.*, **23**, 203.
Terenin, A., and Solonitzin, Y., 1959, *Disc. Faraday Soc.*, **28**, 28.
Wagner, C., and Hauffe, K., 1938, *Z. Elektrochem.*, **44**, 172.
———, and Zimens, K. W., 1947, *Acta Chem. Scand.*, **1**, 547.
———, *Z. Elektrochem.*, in press.
Wood, R. J., and Wise, H., 1958, *J. Chem. Phys.*, **29**, 1416.
Wortman, R., Gomer, R., and Lundy, R., 1957, *J. Chem. Phys.*, **27**, 1099.
Young, F. W., Jr., 1960, *Acta Met.*, **8**, 117.

9. LIQUID ELECTRODE REACTIONS AND CORROSION

9.1. Electrode Processes

Electrochemical changes, either induced or spontaneous, require transfer of a charged particle across a solid-liquid interface. The nature of the particle in any given system depends largely on bulk properties but the transfer rate is almost wholly determined by surface effects. The effect of imperfections, composition, etc., on solid reactions is dealt with elsewhere in this chapter (sec. 4). Here attention is devoted primarily to questions of rate and mechanism, both in electrode processes and in corrosion reactions.

There is now good insight into the kinetics of many kinds of electrode processes in aqueous media at about 300°K. The emphasis in recent years has been on gaining better understanding of the mechanisms of electrode reactions proceeding at finite rates. For processes in which there is a net current flow from electrode to solution, or vice versa, the potential of the reaction deviates from its reversible value to an extent which depends on the system. The difference between the reversible and the measured value, *i.e.*, the overvoltage (or overpotential), is a measure of the irreversibility of the system and is a function of the current density. The phenomenon is often referred to, in general, as electrode polarization. Overvoltage may result from any combination of three causes: (a) ohmic resistance, (b) change in concentration of the electro-

chemically active species by virtue of the electrode reaction, and (c) a slow step in the reaction sequence itself. The first two of these, resistance and concentration overvoltage, usually can be eliminated or taken into account. The last, activation overvoltage, is of greatest interest in the elucidation of mechanisms.

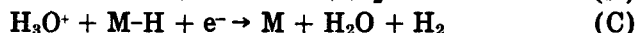
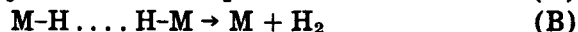
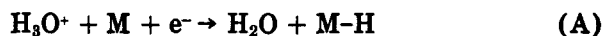
Activation overvoltage is generally most evident in processes in which there are phase changes, and is especially notable in some reactions involving gases. The hydrogen electrode reaction has been predominant in the study of activation overvoltage and the current relatively clear view of this whole problem stems in large part from the insight gained from observations on electrochemical reduction to hydrogen gas. For this reaction, Tafel (1905), half a century ago, established empirically the important fact that the overvoltage, η , is linearly dependent on $\ln i$, where i is the current density, usually in amps/cm²:

$$\eta = a + b \ln i. \quad (9.1)$$

The constants of this relationship have since been the subject of extensive theoretical treatments. Most of these have provided independent methods for evaluating the so-called Tafel slope, b , and some have accounted for the intercept a . The literature on this reaction is voluminous but because of some experimental difficulties which were not always appreciated some of it is misleading. Certain observations, however, may be accepted as being firm (Hickling, 1949; Grahame, 1955R). A striking feature is the marked dependence of η on the nature of the electrode material: thus, metals of low melting point usually show high overvoltage, metals which are catalysts in hydrogenation processes usually show low overvoltage, etc. Overvoltage builds up with time, depending on the electrode's attainment of steady state. It appears to be independent of pH in dilute acid solutions but decreases with increasing acidity in more concentrated solutions. Fewer data are available for alkaline solutions, but Eq. (9.1) seems to be obeyed, although b is considerably greater than in acid solutions. The addition of other materials to the solution affects the overvoltage. For example, HgCl₂ or CS₂ cause an increase in overvoltage at Pt or Ni cathodes, while alkaloids of the quinoline group decrease it at Hg cathodes. The effect of the solvent is still largely unknown although maxima have been observed in mixed aqueous-non-aqueous systems. Hydrogen overvoltage exhibits a negative temperature coefficient of 2 to 3 mv/°C for high overvoltage metals and smaller coefficients for low overvoltage metals. Finally, deuterium overvoltage is larger than that for hydrogen under the

same conditions and the ratio differs for different cathodes, as indicated by different separation factors.

Good accounts of current theories of hydrogen overvoltage, based on attempts to determine the slow step in the reaction sequence, exist (Bockris, 1945R; Frumkin et al., 1952R). It is clear that hydrogen gas evolution at a cathode may stem from the reduction of a hydrated proton, a water molecule, some unstable intermediate, or some combination of these. The theories are almost wholly concerned with the process in acid solution and therefore usually consider the hydrated proton as the starting material. Three reactions in particular are believed to be involved, any one of which may be rate determining:



The slow discharge theory (A) was treated from different viewpoints by Erdey-Grúz and Volmer (1930) and by Gurney (1931) and developed further by others (Frumkin, 1937, 1943, 1947; Butler, 1936; Horiuti and Polanyi, 1935). Fundamentally, it provides for an energy barrier which requires an activation energy, w , at the reversible potential. The activation energy follows the Maxwell distribution law and is a function of potential. At the reversible potential the rate of hydrogen discharge is equal to the rate of oxidation of hydrogen gas and each is related to w exponentially. If the electrode is made the cathode of an electrolytic cell and the potential on it is altered by an amount η , the activation energy is given by $w - \alpha \eta F$ where $0 < \alpha < 1$ and F is the Faraday. If the cathode is sufficiently polarized so that the rate of the reverse (oxidation) reaction can be neglected, the reduction rate is now given in terms of current density by

$$i = Fk_1a_+ e^{-(w - \alpha \eta F)/RT}. \quad (9.2)$$

This reduces to a form of the Tafel equation which, written non-dimensionally, is

$$\eta = \frac{RT}{\alpha F} \ln \frac{i}{i_0}. \quad (9.3)$$

Here i_0 is the exchange current density, i.e., the rates of the opposing reactions while the electrode is at the reversible potential. Equations (9.2) and (9.3) show that the Tafel plot can be used to obtain α and i_0 . The slope $[2.3RT/\alpha F]$ is often approximately 0.12 and therefore at 25°C $\alpha \approx 0.5$. The value of w is obtainable from the temperature dependence of $\log i$. The evaluation of α , i_0 , and w has been the main preoccupation of electrochemical

kinetic studies. The value of α is generally believed to be significant for the slow discharge mechanism (Horiuti and Ikusima, 1939; Horiuti, 1949).

This theory gives results which, in some respects, are in satisfactory agreement with experiment, particularly for the mercury cathode. However, it provides no real explanation of the marked dependence of overvoltage on cathode material. Qualitatively, it can be said that the energy of activation, and hence the overvoltage, is lowered when a metal is a good adsorbent for hydrogen atoms, but no detailed treatment of this point is available. Also, the theory provides no explanation for the time dependence, for the effect of solvent, or for catalytic poisons. The nature of the activated step is not specified and may involve either the actual discharge of the ion or its dehydration¹.

The catalytic mechanism (B) ascribes overvoltage to the accumulation of hydrogen at the cathode, with the rate of recombination as the controlling step. It can be shown thermodynamically that the accumulation of atomic H at the cathode should give rise to potentials more negative than the reversible potential. Thus, metals that are good catalysts for the combination of H atoms ought to give lower overvoltage, and those with poor catalytic activity higher overvoltage, as is the case. Rapid diffusion of cathodic hydrogen through metals, action of catalytic poisons, and the transference of overpotential from the polarized to the unpolarized side of iron and palladium diaphragms separating two electrolytes, are cited in support of this theory. From the kinetic point of view, if it is assumed that the potential is proportional to the logarithm of the adsorbed H atom concentration, the dependence of overvoltage on current density is given by an expression similar to Eq. (9.2) or (9.3), but with $\alpha = 2$, making $b \cong 0.03$ instead of ~ 0.12 , as frequently observed experimentally and as predicted by the slow discharge mechanism. It has been suggested that there is no theoretical basis for considering the potential as directly dependent upon the logarithm of the adsorbed H atom concentration. Instead, the pressure of free atomic H in equilibrium with the adsorbed hydrogen should be considered. For

¹ A theory has been proposed in which the transfer of a proton from a water molecule in solution to an adsorbed water molecule on the cathode is the slow step. By means of the absolute reaction rate theory, an equation is derived in which $\alpha = 0.5$. The variation of overvoltage with cathode material is explained in terms of M-H bond strength. Experimentally the change of solvent has a much smaller effect than predicted by this theory (Eyring, et al., 1939; Kimball, et al., 1941; Butler, 1941).

poor adsorbents like mercury, or for surfaces approaching saturation, this gives much larger values of b .

The electrochemical mechanism (C) was proposed as the rate-determining step by Horiuti and his co-workers (1935) and provides fair quantitative agreement with experimental observations. Assuming (C) to be a fast reaction and to be of importance at surfaces largely covered with adsorbed hydrogen atoms, and making some assumptions as to the relative participation of reactions (B) and (C) in the removal of adsorbed hydrogen, it is possible to account for: (a) the negative deviations from the Tafel equation at high current density; (b) the approach to constant values of overpotential; and (c) the observed separation factors of hydrogen and deuterium.

If more than one rate-determining step exhibits about the same Tafel slope, it is not possible to ascertain the controlling step from α (or b) alone and additional criteria must be established. For the purpose, the stoichiometric number, γ , for the reaction has been used (Makrides, 1957). This is the number of times the rate-determining step occurs when the over-all reaction occurs once. Assuming that α is known and is less than one, the usefulness of the stoichiometric number lies in its ability to distinguish between (B) and (C) as the follow-up step. However, validity of the Tafel relation near the reversible potential must be established, since γ is obtained from the exchange current density by extrapolation to zero overvoltage. It has been suggested, from the slow-discharge theory, that the overvoltage is directly proportional to current density as $i \rightarrow i_0$. Also, γ is difficult to obtain because of experimental problems in measuring overvoltages at the very low current densities approaching the usual exchange value.

It is entirely possible that the mechanism may be different for metals, or that the overall rate is controlled by more than one step, or that, for a given metal, the rate-controlling step varies depending on current density and surface coverage, among other factors. A theoretical treatment has recently been attempted for a system in which there are two probable rate-determining steps operating simultaneously (Bockris and Mauser, 1959). The present position appears to be that, at high overpotential cathodes (mercury in acid solutions), the slow discharge mechanism is in closest agreement with the experimental observations; at low overvoltage cathodes (Pt, Pd, Au, W, Mo) the rate-determining step is the formation of molecular H_2 from adsorbed H atoms. There is no general theory, applicable to both high and low overvoltage materials.

An entirely different approach to the problem of classifying the mechanism of electrode reactions is that of Vetter (1953, 1955) and of Gerischer (1954). They consider typical redox reactions not involving gases and concern themselves with determining the electrochemical reaction order. The current (i.e., rate of electrode reaction) is studied as a function of potential, rate of stirring, and concentrations of various substances present in the solution. As the potential increases, there is always a limiting current at a given rate of stirring, and this generally increases with the rate of stirring. In such a case, the limiting rate is diffusion-controlled, and the magnitude of the limiting current as a function of the concentrations of the molecular species present usually tells what substance is limiting the rate by its diffusion to the electrode. If the limiting current is independent of the rate of stirring, it indicates the presence of a slow precursor reaction, one of whose products is needed for the electrochemical reaction. If the limiting rate is decreased by stirring, it indicates that one of the products of the electrochemical reaction is needed for a slow precursor reaction.

Experimentally, there are still considerable difficulties. It is possible now to purify solutions satisfactorily and to provide circuitry for fast, sensitive, and accurate electrical measurements. The use of the potentiostat (Gerischer and Stauback, 1957) with fast rise times has been especially helpful. However, the solid surface of the electrodes still poses some unanswered questions. Reproducibility leaves a great deal to be desired and nothing quantitative can be said about the influence of the surface on reactions, either when acting as an electron membrane or as part of a redox system. Although there is now sound knowledge of the kinetics of hydrogen evolution on some metals, e.g., Hg, Pt, Ag, and Ni from aqueous solutions at 300°K, the information is restricted to these few metals in acid or alkaline solutions. Good experimental information is not yet available for iron and other metals of interest, for other media, or for systems at considerably elevated temperatures. One major problem which has been noted only in passing in the past, but which now appears to be the subject of much more interest, is that of the effect of adsorption of non-participating ions on electrode reaction kinetics. For instance, the electro-reduction of $S_2O_8^{2-}$ is quite sensitive to the presence of non-reducible cations, as is the reduction of hydrogen ion in the presence of iodide ion (Frumkin, 1959).

The kinetics of electrodeposition of metals are less well understood in spite of a long history and great technological importance.

These systems pose the problem of the formation of a new phase with the necessity for nucleation and the complication of continual surface change, and these factors are undoubtedly of major importance in the overall kinetics. Additional difficulties are caused by the high exchange currents which limit the usefulness of d.c. methods, since large ohmic and concentration overvoltages are introduced which obscure the relatively small activation overvoltages. The particular problem here is to establish how and in what order the steps of cation desolvation, transfer to a lattice site, and electron transfer take place. A good description of how the charge transfer process might take place on the basis of complex ion theory (Lyons, 1954), and more conventional theoretical and experimental treatments of the deposition of metal (Conway and Bockris, 1958; Mehl and Bockris, 1957; Gerischer, 1958; Vermilyea, 1958) are indicative of the present approaches. The view that electron transfer takes place without the electron leaving the metal lattice has been challenged (Randles and Somerton, 1952). The life of an electron in transit between the electrode and the redox couple and the role of field emission in the electrical double layer is a general problem in electrode reactions.

Perhaps the most significant recent advances in electrode processes are those associated with work on the electrical double layer (edl) and adsorption at polarized electrodes. Within a certain potential range, the metal-solution interface of a polarized electrode behaves somewhat like an electrical condenser because of the asymmetrical distribution of charged components of the two phases across the interface. It is possible, in principle, to measure the capacity of this double layer and thereby to investigate the magnitude and distribution of charges. This is important in electrochemical kinetics because the structure of the double layer influences the effective difference of potential which favors or hinders electrochemical reactions. Further, the effective concentration of ions at the interface is different from that in the bulk and the rate of the electrode process is affected accordingly. Capacity measurements also provide a convenient tool for studying adsorption-desorption of neutral molecules at electrode surfaces.

Experimental approaches involve the use of either an impedance bridge or the method of charging curves (Parsons, 1954R). A modification of the latter involves the use of square-wave signals (Popat and Hackerman, 1958). The most extensive investigations of the structure of the electrical double layer, especially at mercury electrodes, were carried out by Grahame (1958R) and by Frumkin (1959). Mathematical treatment is available for the nearly ideal

system in which there is no net transfer of charge and in which specific adsorption can be neglected (Grahame and Soderberg, 1954; McDonald, 1957). Studies are now being made on solid metal electrodes using both the bridge method and charging curve techniques (Popat and Hackerman, 1958). Proper evaluation of this work is not yet possible because of the difficulties of separating capacitative components from the faradic impedance for solid metal electrodes (Bockris and Conway, 1958).

Two other problems which still must be dealt with concern the kinetics of formation and of decay of the double layer, and the state of the double layer under conditions of net charge transfer. Each should be amenable to procedures using high frequency input signals which would permit measuring the decay of the double layer in the first instance and the resistive component of the working electrode in the second. In spite of these problems, significant improvements in the understanding of the structure of the double layer have been attained. Thus, for example, it is well established that certain anions are chemisorbed and are dehydrated before their discharge, while others are not specifically absorbed (Popat and Hackerman, 1958; McDonald, 1957). As already noted, it has been shown that certain cations are specifically adsorbed at negatively charged mercury surfaces and that they influence the electroreduction process (Frumkin, 1959). Kinetics of adsorption of neutral organic molecules at electrode surfaces and their influence on electrode processes (especially on hydrogen overvoltage) have been studied intensively (Delahay, 1958, 1957R; Breyer, 1956R; Frumkin, 1957) but good generalizations are not yet possible.

Measurement of edl capacities of semiconductor-solution interfaces may be of use in clarifying current problems of abnormal behavior of semiconductors. The capacity-potential curves for Ge have been examined with some success (Bohnenkamp and Engell, 1957). A markedly lower sensitivity to potential was found than for metals. This is probably because of the contribution of the double layer within the semiconductor. There accrues both an advantage and a disadvantage in this: the former because of insight which might be gained in regard to surface states; the latter because a clear assessment of adsorption from the solution side becomes less likely.

There are an abundance of experimental investigations relating to anodic processes in which the electrode material is oxidized, but success in terms of better understanding has been only moderate at best. The information now available concerning electrochemical passivation of pure metals and alloys is quite

reliable. It has been established that considerably less than 0.01 coulomb/cm² is needed to passivate iron, perhaps as little as 300 μ c/cm² being enough. Also, it is clear that only a small fraction of this amount is needed to maintain passivity. It is agreed that the phenomenon must be the result of a film containing some of the metal ions, but probably rich in oxygen, and not thicker than 20Å or so. However, there is a wide divergence of opinion as to the method of formation of this layer and as to how the observed properties are related to its structure. A recent symposium on this subject summarizes the various positions very well (Z. Elektrochem, 1958; Stern, 1957; Gerischer, 1958).

Electropolishing by anodic treatment has progressed beyond the purely experimental stage and the anodic behavior of metals capable of forming thin oxide films has been extensively studied (Jacquet, 1956; Dewald, 1957). The problems are largely those of ion transport through the film under large potential gradients. There is some relation to problems of passivity, as for example, the questions of film density and how best to describe the dielectric character of films. The same situation does not hold for the anodic formation of rectifying oxides or for electromachining by anodic methods—a rather recent innovation (Uhlir, 1955). These areas require extensive additional knowledge of current density distributions and of material transport. In general, the problems of diffusion-controlled electrode reactions are fairly well understood for nonturbulent flow systems. However, only poorly reproducible information is available for systems in which there is turbulence. Similarly, current distribution in uncomplicated electrolytic conduction paths with low currents and small areas can be handled competently. However, not much has even been attempted in predicting behavior with large currents and current densities (e.g., > 10 amps and > 10² amp/cm²). In this connection, current-voltage responses of electrochemical cells have been made to perform some of the functions of electronic components. Development in this interesting area is slowed by the problems noted above (Hurd and Lane, 1957).

A few other areas of electrochemistry have been the subject of renewed interest. Fuel cells of the classic kind or of a regenerative character have been subject to intense technical and minor scientific study (Bacon and Forrest, 1955R; Rideal, 1958; McKee and Adams, 1944; Yeager, 1958). Many variations in primary cells have been devised by application of experience plus a small leavening of new information. Of most interest, perhaps, are cells using a solid electrolyte.

There have been no major scientific advances in electrowinning or in electroplating but technical improvements have been achieved here too. Although, for many years much use has been made of electrochemical processes in fused baths, there has been little improvement in understanding of the electrode processes. A few careful laboratory studies have now been made (Laitinen and Graur, 1957, 1958) by methods similar to those used for electrical double layer and electrochemical kinetic studies in aqueous solutions, but this still remains an almost completely open field for basic research.

The same can be said about the application of irreversible thermodynamics to electrochemical processes. Some serious work of this kind has been undertaken (Van Rysselberghe, 1955R), but much remains to be done. Finally, an active start has been made in studies of the electrochemistry of semiconductors (Bohnenkamp and Engell, 1957; Dewald, 1959R; Gatos, 1959R) and in the electrochemistry of membranes (Sollner, 1950) and these fields are likely to progress rapidly.

9.2. Metallic Corrosion

Research in the corrosion of metals now involves increasing proportions of purely scientific inquiry, although this kind of work still represents only a small percentage of the total effort. Apart from the use of modern solid state theory, e.g., of interactions between the environment and crystal imperfections (Petrovich, 1959), the most notable advance is in the quantitative application of kinetic electrochemistry. Stern (1957) has shown how to use the anodic polarization diagram for passivable metals to predict chemical passivity by an oxidizing agent, or indeed by any form of "anodic protection." This certainly has led to a clearer understanding of the kinetics of uniform corrosion processes in conducting solutions, but does not indicate the nature of those sites at which the electrode reactions occur. Non-uniform corrosion effects in which the locations of the electrode reactions can be predicted, as with crevices, are also understood. However, factors which promote pitting, stress-corrosion cracking and other localized corrosion phenomena are recognized but obviously not understood (see below). Electrochemical principles provide the background for major engineering achievements in cathodic protection. Here again, ignorance of current distribution in strange geometries, at high current values, or under other non-ordinary conditions generally is reflected in an inability to cope with specific problems. Studies on anodic polarization and electrochemical passivation

There have been no major scientific advances in electrowinning or in electroplating but technical improvements have been achieved here too. Although, for many years much use has been made of electrochemical processes in fused baths, there has been little improvement in understanding of the electrode processes. A few careful laboratory studies have now been made (Laitinen and Graur, 1957, 1958) by methods similar to those used for electrical double layer and electrochemical kinetic studies in aqueous solutions, but this still remains an almost completely open field for basic research.

The same can be said about the application of irreversible thermodynamics to electrochemical processes. Some serious work of this kind has been undertaken (Van Rysseberghe, 1955R), but much remains to be done. Finally, an active start has been made in studies of the electrochemistry of semiconductors (Bohnenkamp and Engell, 1957; Dewald, 1959R; Gatos, 1959R) and in the electrochemistry of membranes (Sollner, 1950) and these fields are likely to progress rapidly.

9.2. Metallic Corrosion

Research in the corrosion of metals now involves increasing proportions of purely scientific inquiry, although this kind of work still represents only a small percentage of the total effort. Apart from the use of modern solid state theory, e.g., of interactions between the environment and crystal imperfections (Petrovich, 1959), the most notable advance is in the quantitative application of kinetic electrochemistry. Stern (1957) has shown how to use the anodic polarization diagram for passivable metals to predict chemical passivity by an oxidizing agent, or indeed by any form of "anodic protection." This certainly has led to a clearer understanding of the kinetics of uniform corrosion processes in conducting solutions, but does not indicate the nature of those sites at which the electrode reactions occur. Non-uniform corrosion effects in which the locations of the electrode reactions can be predicted, as with crevices, are also understood. However, factors which promote pitting, stress-corrosion cracking and other localized corrosion phenomena are recognized but obviously not understood (see below). Electrochemical principles provide the background for major engineering achievements in cathodic protection. Here again, ignorance of current distribution in strange geometries, at high current values, or under other non-ordinary conditions generally is reflected in an inability to cope with specific problems. Studies on anodic polarization and electrochemical passivation

of metals have led to the concept of anodic protection (Edeleanu, 1954). Much remains to be understood here, especially because of the possibility of accelerated dissolution of alloys in the "transpassive" region. This may be a particularly serious problem for metals under stress, because it has recently been found (Hoar and West, 1958; Hines and Hoar, 1958; Gerischer, 1957) that high current density anodic dissolution of austenitic stainless steel can be effected with little polarization provided a tensile stress is maintained normal to the dissolving surface. Nevertheless, the possibility of reducing or eliminating pitting by anodic polarization deserves considerable effort.

The area of most importance now and for some time to come is that of the combined effect of mechanical stress and chemical reaction on metals (Robertson, 1956R; Rhodin, 1959R). Considering the case of stress-corrosion cracking alone, it is apparent that there are presently at least two major schools of thought. One believes that the major factor is that of selective attack with the level of stress only of secondary importance. There is no uniformity of opinion on the nature of selective attack but some agreement is likely to be developed in the near future. The other is convinced that the attack is not especially selective but that there are local areas of embrittlement in ductile metal which rupture when chemically attacked (sec. 4.9). Questions as to why the phenomenon has occurred only with certain alloys, and why a given alloy suffers such damage only in some solutions and not in others at the same temperature and with the same major constituents are recognized but cannot be answered yet. The most interesting prospects at present are those which are being developed via advanced deformation theory (Robertson, in press) and the finding mentioned earlier as to the almost unpolarized anodic dissolution of a metal under stress.

The effect of stress on the potential of reacting metals has often been attributed to changes in thermodynamic properties. However, it is quite clear that these effects arise from changes in the mixed reaction potential imposed on the metal. This is supported by the fact that they are of the order of 10–100 millivolts whereas calculated thermodynamic changes are $\sim 10^{-5}$ mv (Rhodin, 1959R). Experiments in which a potential transient is obtained on imposing a momentary stress on a wire (Funk et al., 1957) also support the argument that it is the mixed potential which changes.

There is, at present, a rather large backlog of experimental information on the effects of various kinds of radiation on the corrosion of some metals under rather severe conditions. It is clear that

radiation sometimes causes chemical changes in the environment which influence the reaction rate. Other effects are on barrier layers formed on the metal surfaces. Empirical equations relating corrosion rate to fission power density have been helpful when it can be assumed that fissionable isotopes are incorporated into the surface layers (Jenks, 1956). There seems to be no specific effect on the metal reactivity, although it is very likely that radiation should cause a reduction in overvoltage.

Other items of more than passing interest include kinetic studies with single crystals of some metals (Kruger, 1958); careful morphological studies of solid reaction products (Feitknecht, 1958); relationship between molecular structure and corrosion inhibition by organic compounds (Kaesche and Hackerman, 1958); and passivation by chemical means (Stern, 1957; Cartledge, 1958). These studies have only begun to lead to firm conclusions and to new ideas. For instance, it now appears reasonable to approach the question of predicting inhibitor efficiency by determining the effect on Tafel slopes, exchange currents, "critical" potentials and currents (Kaesche and Hackerman, 1958).

In summary, a few of the major problems deserving continued research are: the true relationship between the nature of the solid and electrochemical processes involving dissolution or formation of elemental metals; the question of specificity of environment and of alloying in stress corrosion cracking; the lack of experimental knowledge of the effects of high temperature, radiation, and possibly very high pressures, individually and collectively, on corrosion and electrochemical processes; the problems of surface preparation for providing reproducible results; a good general theory for all electrode polarization phenomena, if that is possible. Such theoretical development requires more precise experimental information on the time dependence of polarization, solvent effects, and on sorption phenomena in general.

9.3. References

9.3.1. Review and General Articles

- Bacon, F. T., and Forrest, J. S., 1955, 5th World Power Conf., Paper 119 K/4.
Bockris, J. O'M., 1954, *Modern Aspects of Electrochemistry*, Chapter 4, Academic Press, Inc., New York.
Breyer, B., 1956, *Revs. Pure and Applied Chem.*, **6**, 249.
Delahay, P., 1957, *Ann. Revs. Phys. Chem.*, **8**, 229.
Dewald, J. F., 1959, *Semiconductors*, Chapter 17, ed. by N. B. Hannay, ACS Monograph No. 140, Reinhold Publishing Co., New York.
Frumkin, A. N., Bagotskii, V. S., Iofa, Z. A., and Kabanov, B. N., 1952, *Kinetics of Electrode Processes*, Moscow University Press, Moscow.

- Gatos, H. C., editor, 1959, *The Surface Chemistry of Metals and Semiconductors*, New York, John Wiley and Sons, Inc.
- Grahame, D. C., 1955, *Ann. Rev. Phys. Chem.*, **6**, 338.
- , 1958, *Ann. Revs. Phys. Chem.*, **6**, 337.
- Parsons, R., 1954, *Modern Aspects of Electrochemistry*, Chapter 3, Academic Press, Inc., New York.
- Proceedings, 9th Meeting, CITCE, 1959, pp. 425–462, Butterworth's Scientific Publications, London.
- Rhodin, T. N., editor, 1959, *Physical Metallurgy of Stress Corrosion Fracture*, New York, Interscience Publishers.
- Robertson, W. D., editor, 1956, *Symposium on Stress Corrosion Cracking*, New York, John Wiley and Sons, Inc.
- Symposium on Electrode Processes, May 1959, 115th Meeting of the Electrochemical Society, Philadelphia, Pa.
- Van Rysselberghe, P., 1955, *Electrochemical Affinity*, Hermann et Cie., Paris.

9.3.2. Original Articles

- Adams, A. M., 1949, *Fuel*, **28**, 1.
- Bockris, J. O'M., and Conway, B. E., 1958, *J. Chem. Phys.*, **28**, 707.
- , and Mauser, H., 1959, *Canadian J. Chem.*, **37**, 475.
- Bohnenkamp, K., and Engell, H. J., 1957, *Z. Elektrochem.*, **61**, 1184.
- Butler, J. A. V., 1936, *Proc. Roy. Soc. (London)*, **A157**, 423.
- , 1941, *J. Chem. Phys.*, **9**, 279.
- Cartledge, G. H., 1958, *Z. Elektrochem.*, **67**, 684.
- Conway, B. E., and Bockris, J. O'M., 1958, *Proc. Roy. Soc. (London)*, **248A**, 394.
- Delahay, P., 1958, *Rec. Chem. Prog.*, **19**, 83.
- Dewald, J. F., 1957, *J. Phys. Chem. Solids*, **2**, 55.
- Edeleanu, C., 1954, *Metallurgia*, **50**, 113.
- Erdey-Grúz, T., and Volmer, M., 1930, *Z. physik. Chem.*, **150**, 213.
- Eyring, H., Glasstone, S., and Laidler, K. J., 1939, *J. Chem. Phys.*, **7**, 1053.
- Feitknecht, W., 1958, *Z. Elektrochem.*, **62**, 795.
- Frumkin, A. N., 1937, *Acta Physicochim. USSR*, **7**, 475; 1943, *ibid.*, **18**, 23.
- , 1947, *Disc. Faraday Soc.*, **1**, 57.
- , and Florianovich, G. M., 1955, *J. Phys. Chem. USSR*, **29**, 1827.
- , 1957, *Nova Acta Leopoldina*, **19**, 132.
- , 1959, *Trans. Faraday Soc.*, **55**, 156.
- Funk, A. G., Giddings, J. C., Christensen, C. F., and Eyring, H., 1957, *Proc. Natl. Acad. Soc., U. S.*, **43**, 421.
- Gerischer, H., 1954, *Z. physik. Chem. (N.F.)*, **2**, 79.
- , and Stauback, K. E., 1957, *Z. Elektrochem.*, **61**, 789.
- , 1958, *Z. Elektrochem.*, **62**, 256.
- , 1958, *Z. anorg. Chem.*, **70**, 285.
- Grahame, D. C., and Soderberg, B. A., 1954, *J. Chem. Phys.*, **22**, 449.
- Gurney, R. W., 1931, *Proc. Roy. Soc. (London)*, **A134**, 137.
- Hickling, A., 1949, *Quart. Rev.*, **3**, 95.
- Hines, J. G., and Hoar, T. P., 1958, *J. Appl. Chem. (London)*, **8**, 764.
- Hoar, T. P., and West, J. M., 1958, *Nature*, **181**, 835.
- Horiuti, Z., and Polanyi, M., 1935, *Acta Physicochim. USSR*, **2**, 505.
- , and Ikusima, M., 1939, *Proc. Imp. Acad. (Tokio)*, **15**, 39.
- , 1949, *J. Res. Inst. Catalysis*, **1**, 8.
- Hurd, R. M., and Lane, R. N., 1957, *J. Electrochem. Soc.*, **104**, 727.

- Jacquet, P. A., 1956, *Metall. Rev.*, **1**, 157.
- Jenks, G. H., 1956, ORNL-2222 to U. S. Atomic Energy Commission.
- Kaesche, H., and Hackerman, N., 1958, *J. Electrochem. Soc.*, **105**, 191.
- Kimball, G. E., Glasstone, S., and Glassner, A., 1941, *J. Chem. Phys.*, **9**, 91.
- Kruger, J., 1958, *Ind. Eng. Chem.*, **50** (3), 55A.
- Laitinen, H. A., and Graur, H. C., 1957, *J. Electrochem. Soc.*, **104**, 730; 1958, **105**, 444.
- Lyons, E. H., 1954, *J. Electrochem. Soc.*, **101**, 363, 410.
- Makrides, A. C., and Hackerman, N., 1955, *Ind. Eng. Chem.*, **47**, 1773.
- , 1957, *J. Electrochem. Soc.*, **104**, 677.
- McDonald, J. R., 1957, *J. Chem. Phys.*
- McKee, J. H., and Adams, A. M., 1949, *Fuel*, **28**, 1.
- Mehl, W., and Bockris, J. O'M., 1957, *J. Chem. Phys.*, **27**, 818.
- Petrovitch, A. L., 1959, *Ind. Eng. Chem.*, **51**, 74A.
- Popat, P. V., and Hackerman, N., 1958, *J. Phys. Chem.*, **62**, 1198.
- Randles, J. E. B., and Somerton, K. W., 1952, *Trans. Faraday Soc.*, **48**, 951.
- Rideal, E. K., 1958, *Z. Elektrochem.*, **62**, 324.
- Robertson, W. D., and Tetelman, A. S., in *Strengthening Mechanisms in Solids*, ASM (in press).
- Sollner, K., 1950, *J. Electrochem. Soc.*, **97**, 139C.
- Stern, M., 1957, *J. Electrochem. Soc.*, **104**, 645.
- Tafel, J., 1905, *Z. physik. Chem.*, **50**, 641.
- Trans. Electrochem. Soc.*, 1939, **76**, 145.
- Uhlir, A., Jr., 1955, *Rev. Sci. Instr.*, **26**.
- Vermilyea, D. A., 1958, *J. Chem. Phys.*, **28**, 717.
- Vetter, K. J., 1952, *Z. Naturforsch.*, **7a**, 328; 1953, **8a**, 823.
- , 1955, *Z. Elektrochem.*, **59**, 596.
- Yeager, E., Dec. 1958, Tech. Rep. 10 to Office of Naval Research (Proj. NR 359-277).
- Z. Elektrochem*, 1958, **62**, Nos. 6-7.

10. A SUMMARY OF UNSOLVED PROBLEMS

This section summarizes briefly those problems which are common to all or most aspects of the study of surfaces.

10.1. Structure

There are three main problems connected with structure, namely, determination, control, and assessment of its effects. It is possible today to characterize macroscopic specimens quite well with respect to gross but not with respect to atomistic structure. Microspecimens, e.g., field emission tips, can be characterized even on the atomic scale. At present there is very little knowledge of the detailed structure of composite surfaces, interfaces, ad-layers and so on.

Macroscopically smooth and clean surfaces can be prepared today for large specimens of high melting metals, and atomically smooth surfaces of these materials can be obtained on a micro-scale. This is not true for most other substances. It is nevertheless

deplorable that much work is still being done on surfaces less carefully prepared and characterized than presently possible.

Information on the effect of microstructure on all surface phenomena is completely qualitative at present, and largely confined to rather naive physical models. Much more detailed and sophisticated approaches are needed.

10.2. Energetics

There is little reliable information on surface and interfacial energies and tensions, and almost no information on their variation with orientation or composition. Heats of adsorption are known, with some degree of reliability, for only a few gases on a few metals, but even here variations with coverage seem to change erratically from investigation to investigation. This is undoubtedly to be associated with insufficiently cleaned and inhomogeneous surfaces. Work function data are reliable for only a few metals and perhaps one semiconductor; the variation with orientation is known with precision for only one metal. Information on the effect of alloying and adsorption is greatly needed.

On the theoretical side, the calculation of surface energies is in a most rudimentary state, and attempts to take structure into account seem rather beyond realistic aims at the moment. Almost nothing of a theoretical nature is known about the details of ad-bonding. Empirical correlations of surface diffusion activation energies and heats of binding have been made, but theoretical treatments are wholly lacking. Little theoretical knowledge exists about the energetics of grain boundaries and impurity adsorption at boundaries.

10.3. Reactivity

Details concerning the formation of very thin films (0–100Å) are still unknown. The influence of imperfections is very poorly understood, experimentally or theoretically. The kinetics of electron transfer in adsorption and in electrolytic processes are poorly understood.

10.4. Conclusions

Some phases of surface work, notably those involving microspecimens of clean metals, have undergone a revolution in the last 10 years. Others wait to be liberated from a host of environmental and semi-accidental variables which place them in the category of art rather than science. Theoretical developments have not kept pace with experimental progress in any phase of surface study.

Part X

STRUCTURE AND PROPERTIES OF LIQUIDS

Chairman

Prof. BRUCE CHALMERS
Harvard University
Cambridge, Mass.

Members

Dr. B. R. T. FROST
Atomic Energy Research
Establishment
Harwell
Didcot, Berks, England

Prof. ROBERT B. GORDON
Yale University
New Haven, Conn.

Dr. KENNETH H. JACKSON
Harvard University
Cambridge, Mass.

Prof. GEORGE J. JANZ
Rensselaer Polytechnic Institute
Troy, N. Y.

Prof. PAUL C. MANGELSDORF, Jr.
The University of Chicago
Chicago, Ill.

Prof. STUART A. RICE
The University of Chicago
Chicago, Ill.

NOTE

Perspectives in Materials Research has been in publication far longer than was originally contemplated. In most fields, the changes that occurred during this period were not so great as to render the original text inadequate for the purpose intended. However, this was not the case for the Part on liquids. In order to avoid further delay in the issuance of this volume, it was agreed that the original text of Part X would be deleted. The Office of Naval Research is considering alternative ways of providing a report on perspectives in research on the structure and properties of liquids.

Part XI

EFFECTS OF RADIATION ON MATERIALS

Chairman

Dr. GEORGE H. VINEYARD
*Brookhaven National Laboratory
Upton, Long Island, N. Y.*

Members

Dr. R. S. BARNES
*Atomic Energy Research
Establishment
Harwell
Didcot, Berks, England*

Dr. J. A. BRINKMAN
*Atomics International
Canoga Park, Calif.*

Dr. JAMES H. CRAWFORD, Jr.
*Oak Ridge National Laboratory
Oak Ridge, Tenn.*

Prof. JAMES S. KOEHLER
*University of Illinois
Urbana, Ill.*

CONTENTS

	Page
1. INTRODUCTION	631
1.1. General Remarks	631
1.2. Outline of This Study.....	633
1.3. Acknowledgments	634
2. ASSESSMENT OF THE PRESENT STATUS OF THE FIELD	634
2.1. Models of Radiation Effects: Production Processes	634
2.2. Models of Radiation Effects: Properties of Lattice Defects in Metals	640
2.2.1. Frenkel Defects	641
2.2.2. Vacancies	642
2.2.3. Interstitials	643
2.2.4. Displacement Spikes	644
2.2.5. Depleted Zones	645
2.3. Models of Radiation Effects: Properties of Lattice Defects in Non-metals	646
2.3.1. Semiconductors	646
2.3.2. Insulators	647
2.4. Models of Radiation Effects: Production Processes and Prop- erties of Lattice Defects in Materials Suffering Transmuta- tions	648

	Page
2.5. Existing Experimental Coverage.....	650
2.5.1. Metals and Alloys.....	650
2.5.2. Semiconductors	657
2.5.3. Insulators	662
2.5.4. Materials Suffering Transmutations	665
3. NEEDS AND PROSPECTS.....	667
3.1. Development and Use of New Observational Techniques	667
3.1.1. Etching	667
3.1.2. Thin Films	668
3.1.3. Moiré Patterns	669
3.1.4. Field Emission Microscopy.....	669
3.2. Experiments with Monoenergetic Neutrons and with Very High Energy Neutrons and Other Particles	669
3.3. Invention of New Methods of Distinguishing Between Defects, Particularly Interstitials and Vacancies	671
3.4. Wide Use of Radiation-Enhanced Diffusion	671
3.5. Studies of Transient Damage with Accelerators and Pulsed Reactors	673
3.6. Wider Application of Magnetic Resonance Techniques	673
3.7. Wider Use of the Slow-Neutron-Transmission Technique.....	674
3.8. More Use of Electrons and Gamma-Rays and Intercomparison of Results Obtained with Various Kinds of Radiation	676
3.9. Careful Annealing Studies on Very Pure Irradiated Specimens	677
3.10. Studies of Directional Effects.....	677
3.11. Studies of Chemical Effects in Solids.....	679
3.12. Further Studies of Effects on Elastic Constants.....	681
3.13. Further Study of Effects of Radiation on Plastic Properties	681
3.14. Critical Studies of the Mechanisms of Ordering and Disordering by Radiation.....	682
3.15. More Measurements of Stored Energy.....	683
3.16. Calculations of the Interaction of Point Defects with Dislocations	685
3.17. More Use of Pedigreed Materials.....	686
3.18. More Use of Doped and Pre-Treated Crystals.....	687
3.19. More Extensive Measurements of Threshold Energies for Displacement	687
3.20. Measurements and Calculations of Atom-Atom Interaction Potentials	689
3.21. Direct Calculations Concerning the Displacement Process	691
3.22. Calculations of Energies Associated with Defects	692
3.23. Further Calculations of the Influence of Lattice Defects on Physical Properties	693
3.24. Further Experiments on Growth and Swelling of Fissionable Materials	693
3.25. More Quenching Experiments to Shed Light on Point Defects	696
3.26. More Cold-Work Experiments to Shed Light on Defects	696
3.27. Further Study of the Production of Color Centers in Alkali Halides	697
3.28. Elucidation of the Metamict State: Quartz.....	698

	Page
4. SPUTTERING	699
4.1. Introduction	699
4.2. Experimental Techniques	701
4.3. Established Results	703
4.4. Theories of Sputtering	704
4.5. Needs and Prospects	707
5. REFERENCES	707

EFFECTS OF RADIATION ON MATERIALS

1. INTRODUCTION

1.1. General Remarks

Although the roots of the subject go much further back, radiation effects in solids have been under intensive study for only about 15 years. These 15 years have seen the rapid accumulation of a large number of empirical results of great variety and interest.¹ Studies have been made of the effects of radiation on all of the ordinary physical properties of many solids under a wide variety of conditions. In addition, the effects of radiation on physical processes, such as phase changes, and on some chemical and surface properties have also been explored to a limited extent. The coverage has been far from complete, yet it has been comprehensive enough to expect that fundamental principles would have been established and that a broad understanding of radiation damage phenomena would have been achieved. This is not altogether the case. Because of the great inherent complexity of the phenomena, and because only a small fraction of the experiments have been carried out under sufficiently simplified and well-controlled conditions, understanding is fragmentary and somewhat hypothetical. Simplified atomic models of the radiation damage production processes have been developed. In the best situations, they seem to constitute at least first approximations to the real events; in other instances present models are inadequate or two or more conflicting models exist, each of which fit the experimental facts about equally well. An important collateral activity of recent years has been the serious calculation of the properties of elementary defects in selected materials. Our understanding of radiation effects relies

¹ General reviews are the following: Kinchin and Pease (1955), Seitz and Koehler (1956), Dienes and Vineyard (1957), Harwood, et al. (1958).

heavily on such calculations, and an expansion of these activities in the near future is foreseen.

In this study we have confined our attention largely to fundamental matters and to situations in which an understanding of atomic mechanisms has either been achieved or is in prospect of being achieved. This means that substantially all of the more empirical side of the subject has been ignored, although the importance of rough but exploratory experiments in new areas has not been overlooked.

From our standpoint the study of radiation effects is very closely related to the whole field of defects in crystal lattices. This view springs from the belief that only a small or modest number of fundamentally different lattice defects exist, and that these defects are the "atoms" in terms of which all structure-sensitive properties of solids may be analyzed. These elementary defects can be regarded as embedded in the perfect crystalline lattice, which thus becomes analogous to the real vacuum of ordinary particle physics. The defects have intrinsic properties, and they interact with other defects. When they are present in small concentration their intrinsic properties are manifest; when they are present in higher concentrations their interactions become important. Perhaps this approach is naive—possibly it is altogether wrong. There may be, for example, so many different kinds of defects in even one type of lattice that efforts at analysis in these terms could never be very successful. In any event, it now appears sensible to proceed on the assumption that this approach will be useful in all cases where the concentration of defects is low.

The number of ways in which defects can be introduced into a crystalline lattice actually is rather limited. In addition to those defects originating during growth or formation of a crystal, defects may be introduced by heating followed by quenching, by plastic deformation, and by certain chemical or electrolytic treatments. Finally, defects may be introduced by irradiation. Some defects may be introduced in several ways, others are produced in significant concentrations only by one specific treatment. Irradiation is capable of introducing a rather broad spectrum of point defects and is an extremely powerful and flexible tool in this regard.

Attempts to analyze the irradiated state of a crystal in terms of elemental defects embedded in a perfect matrix may not be at all satisfactory when dealing with a very heavily irradiated substance. Thus, when the defects become too numerous they interact so strongly that their individuality tends to be swamped. Perhaps this "fully irradiated" lattice is best regarded as a new state, possess-

ing a type of perfection of its own. It bears a resemblance to the vitreous state, although it is not necessarily identical with it. The only example which has been investigated extensively is quartz, in which the fully-irradiated state is appreciably different from the vitreous state of the same substance.

As knowledge in the field has increased, and as greater sophistication has been acquired, the character of experimentation has tended to change. Much greater attention has been given to experiments on extremely pure materials, with well-calibrated sources of radiation and low exposures. Especially noteworthy has been the realization that irradiations must be carried out at extremely low temperatures if damage is to be observed in its original state, uncomplicated by annealing processes. Work at liquid helium temperatures is commonplace in a number of laboratories. A commendable tendency to measure a variety of physical properties in the same experiment or on the same substance is now in evidence. Furthermore, the wider use of electron irradiations at moderate energies provides information on a particularly simple species of damage, from which a great deal of fundamental knowledge is being accumulated.

1.2. Outline of This Study

In the following study we attempt first to assess the present status of the field, in its fundamental aspects. To do this it is necessary to discuss both current models of the processes by which matter is disarranged by radiation and current ideas of the structural defects which are produced. We also survey briefly the extent of the coverage of phenomena and materials by simple and well-controlled experiments in which clear-cut interpretations may be expected. Most of this discussion is subdivided according to types of materials—metals, semiconductors and insulators being broadly distinguished, with fissionable materials being assigned to a separate category. We have considered all kinds of energetic radiation in the belief that the differences in response are at least as important as the similarities. Moreover, the effects of radiation on many different physical properties have been considered, but inasmuch as the major part of the experimental and theoretical work in this field has been concerned with a very much smaller set of properties, this state of affairs is naturally reflected in our treatment.

Following the assessment of the present state of the field, pressing needs for the near future are outlined in a section entitled "Needs and Prospects." Problems have been chosen for inclusion which are conspicuously in need of more attention because of their

pivotal position and because conditions seem to be ripe for making significant progress in these areas. Finally, a discussion of the field of sputtering is provided, since sputtering and radiation damage are close cousins and have much in common.

1.3. Acknowledgments

We have been assisted by a number of people in the preparation of this report, and particularly wish to credit G. J. Dienes, who wrote the section dealing with the chemical effects of radiation (sec. 3.11), and G. Wehner and G. Anderson, who contributed the section on sputtering (sec. 4.).

2. ASSESSMENT OF THE PRESENT STATUS OF THE FIELD

2.1. Models of Radiation Effects: Production Processes

The production of defects in a crystal by radiation commences with a primary interaction between the radiation and the solid, which gives rise to excitation or to a knock-on event. The "debris" from such an event bombards the crystal internally, producing further defects. It is convenient to regard this internal bombardment as a secondary stage, and to discuss the primary and secondary stages separately.

The primary interactions are of many forms. When a material is bombarded with high energy particles (electrons, protons, alpha particles, neutrons, etc.) collisions with nuclei always occur and these collisions can be classified as elastic or inelastic. With inelastic collisions, nuclear reactions (excitations or transmutations) take place. Nuclear reactions occur most prominently during ultra-high energy irradiations and in neutron bombardment of certain isotopes; the important case of fissionable materials will be given special consideration subsequently. Extensive data on neutron cross sections of nuclei are now available, although there is need for more data on the angular distributions and the inelastic scattering cross sections for fast neutrons. Calculation of the frequency of primary interactions produced by neutron bombardment is hindered more by uncertainties in reactor fluxes than by uncertainties in the neutron cross sections. It should be noted that the cross-section information comes mostly from experiments, which are rather straightforward in principle, and not from fundamental theory.

For bombardment with heavy charged particles (except particles

having extremely high energies), elastic collisions represent the main interaction, and the differential cross section for scattering processes of this type is well known from the basic theory of Rutherford scattering. For electron bombardment, the relativistic form of the theory of Rutherford scattering must be used, and, where light elements are involved, this allows simple calculations. For heavy elements and electrons with relativistic energies, the theoretical calculations become quite intricate. Specific numerical information in this realm is lacking. Such information would be most useful, particularly in view of the growing importance of experiments involving electron bombardments.

Primary interactions with the electrons in the specimen also occur during bombardment either with charged particles or with electromagnetic radiation. In general this is a very complex matter, understood in a broad and qualitative way, but not in quantitative detail. Electronic effects (excitation and ionization) are of great importance in insulators. They play a minor role in metals, except during gamma bombardment in which knock-on electrons of high energy can produce atomic displacements. Virtually all of the chemical effects of ionizing radiation must be attributed to primary interactions with electrons. To the extent that the energy absorbed is the dominant factor, fair understanding exists; this forms the basis of most radiation dosimetry, but the various modes of disposition of the absorbed energy are not known intimately. The strictly chemical effects of ionizing radiation will be treated only briefly in this study.

It appears then, that the primary stage of the interaction of radiation of various kinds with crystals is understood moderately well, even though many details, particularly those relating to ionization effects, are still obscure. The secondary stage, on the other hand, is much less understood, despite a great deal of effort which has been devoted to it. Major interest attaches to the processes initiated when the primary interaction has produced an energetic knock-on atom or ion with energy in the range 10 to 10^6 ev which is wrenched from its lattice site and plunged forcibly into the surrounding crystal. Basically these are dynamic many-body processes, in which the interaction forces are only approximately known, and whose details are almost hopelessly complex. Full understanding would involve knowing the nature, number, and spatial distribution of all the lattice defects produced for each energy and direction of motion of a primary knock-on.

In practice we settle for much less. General agreement exists about certain features of the secondary stage. Knock-ons of very

high energy are slowed down chiefly by the production of ionization until their energy becomes sufficiently low; thereafter, the interactions which occur are predominantly elastic collisions with atoms of the lattice. The ionization range ends at an energy of roughly 5×10^4 ev for copper, at higher energies for heavier atoms, and lower values for light atoms. The ionization range can logically be included as part of the more general problem of the deceleration of energetic charged particles in matter. Insufficient specific quantitative information is available, but existing theoretical and semi-empirical formulas are adequate for most purposes. The range of elastic collisions, on the other hand, is vital, since it is here that most of the lattice defects are produced. It is also in this range that our knowledge is inadequate (see also sec. 3.20).

The complex events of this stage traditionally have been dealt with by models of two distinct types. Models of the first type emphasize the separateness of the various interactions and assume that a series of secondary knock-on atoms will be produced in distinct, elastic, two-body collisions by the primary knock-on; the secondary knock-ons may then create additional knock-ons by further collisions of the same kind. Such models will be referred to as cascade models. Models of the second type emphasize many-body effects and will be discussed presently.

By assuming that a struck atom must receive more than a critical threshold energy to become displaced, and neglecting the fact that the atoms are arranged in a crystal lattice rather than at random, calculations have been made of the average number of interstitials and vacancies produced in a cascade. Variants of the calculation, assuming hard sphere collisions and soft sphere collisions and that the probability of displacement rises sharply from zero to a value of one at the threshold or rises gradually from zero to one over a range of energies, also have been carried out. The various cascade models are in agreement in predicting that the energy lost in elastic collisions (assuming the initial energy to be well above threshold) is used in creating displaced atoms with about 50 per cent efficiency; that is, the average number of displaced atoms produced per primary knock-on of energy E , assuming all of this energy goes into elastic collisions, is about $E/2E_d$, where E_d is a suitably chosen threshold energy for displacement. Numerous experiments have been designed to test this prediction (Harwood et al., 1958), but in only a small number of instances have all of the desirable experimental precautions been observed (sec. 3.19). With graphite, approximate agreement with theory has been obtained, but with most other materials,

including copper, silver, gold, germanium, alpha aluminum oxide, quartz, and fused silica, the simple cascade theory tends to predict a greater number of displacements than are actually observed after heavy particle bombardment. The theoretical values generally are too large by factors varying from 2 to 20. The explanation of the discrepancy is not known with certainty and may not be the same in all cases.

A few cascade calculations have been carried out for polyatomic substances, in which the differences in atomic masses must be taken into account (Harris, 1955, 1956; Baroody, 1958). If a moving atom makes a head-on elastic collision with an atom of equal mass it transfers its entire energy to the struck atom, but in a collision with an atom of different mass the moving atom never loses all of its energy. Therefore, in polyatomic substances (neglecting differences in atomic size) a tendency for heavy knock-ons to displace only heavy atoms and light knock-ons to displace only light atoms can be predicted. The interaction potentials for the various atom pairs and the thresholds for displacements of different atoms also should be quite different. Such data are important to cascade calculations, and since these quantities have not yet been clearly determined for any polyatomic substance the working out of cascade theory thus far has been on a rather limited and schematic basis. In the comparison of theory with experiment in the above-mentioned examples of silica and aluminum oxide, the assumption of an equivalent monatomic substance has been made. The more detailed investigation of cascade processes in polyatomic substances represents an enormous area for further work.

It can also be predicted that many collisions will occur in which the moving atom knocks an atom out of its lattice site and loses so much energy in the process that it is captured on the vacated site. These "replacement" collisions may be of importance in producing observable effects in compounds and in ordered alloys. There is experimental evidence that the disordering of ordered alloys at low temperatures by heavy-particle irradiation is surprisingly efficient (sec. 3.14), and replacement collisions have been suggested as the explanation. Such a mechanism is highly qualitative, however, since the relevant threshold energies and probabilities of replacements are still matters of speculation.

Greater appreciation of the importance of the regularity of atomic arrangement to the damage progress has been stimulated by the discovery of a "focusing" mechanism for propagation of energy down closely packed lines of atoms (Silsbee, 1957). This

mechanism can be demonstrated easily for an isolated line of hard spheres and would have several important consequences in simple close-packed crystals. In addition to the "easy" directions in which an atom may be displaced by knock-on energies close to threshold, there will be "hard" directions (close-packed directions) in which an atom may not be permanently displaced even when given knock-on energies far above threshold. Energy focused down such a line may produce a displaced atom if it encounters the core of a dislocation, or other existing defect. Furthermore, the focusing effect may assist the "crowdion" configuration of an interstitial (providing it can be formed in the collision process) to possess dynamic stability and thus enable it to propagate away from its point of formation. These and other consequences of focusing by no means have been fully established, but they indicate that there are important complications in the displacement process which are not included in the simple cascade theory.

Cascade models represent one of the two basic approaches to the damage-producing process. At the opposite extreme are those models which emphasize the simultaneous interactions of large numbers of atoms. The various spike models which have been proposed fall in this category. The thermal spike model is based on the assumption that the excited region near the track of a knock-on behaves as if it had been subjected to sudden heating. It is assumed therefore that the excitations have been brought locally to thermal equilibrium, and that the heated region will cool according to the laws of heat conduction. The accuracy of these assumptions has never been tested adequately, and indeed it would be difficult to construct theoretical tests which do not involve large elements of intuition. Qualitative predictions of the effects of thermal spikes are easy to make, involving merely the consequences of rapid localized heating and quenching, but quantitative results seem unreliable. This is because all processes which depend on thermal activation have rates which vary strongly with temperature, so that modest uncertainties in temperature are reflected as large uncertainties in results.

More drastic effects of a many-body nature also have been predicted and have been classified as displacement spikes (Brinkman, 1954; Seeger, 1958). It is envisioned that an entire microscopic region may be disrupted by a knock-on to an extent equivalent to melting and turbulent stirring followed by resolidification into various complicated configurations of damage. Such events should occur toward the end of the range of an energetic knock-on atom, and the effects should be more pronounced in substances of high

atomic weight. The disturbed region may resolidify into a crystalline or into a "liquid-like" or other amorphous configuration and might be of reduced density, because interstitials will be projected out of the region. If interstitials should form as dynamic crowdions, they may be carried far away. A variety of experimental results, particularly in fissionable materials and in noble metals irradiated at low temperatures, could be explained by such effects. Unfortunately, the displacement spike models have not been susceptible to detailed or quantitative theoretical treatment.

It is clear that both the spike and the cascade approaches are over-simplifications. The extent to which either model describes a real radiation damage event depends markedly on the nature of the interaction between a knock-on and other atoms of the lattice. If this interaction has a very short range, that is, if the atomic sizes are effectively small compared with the interatomic distance, collisions will be widely spaced and a cascade type of calculation is justified. If, on the other hand, interactions are of longer range, so that energetic collisions are closely spaced along the track of a knock-on, the spike concepts seem more reasonable. Available evidence indicates that for a knock-on of high energy the effective atomic size is indeed small. As the knock-on slows down, the effective atomic size increases and, at least in common metals, becomes comparable with the interatomic spacing when the knock-on is still moving rather rapidly. It is thus probable that isolated cascade events are initiated along the early stages of the track of an energetic knock-on, and that complicated events, better described by the spike models, occur in the later stages. The vital range of interaction energies for this problem is from about 10 eV to about 10^4 eV, and it is just in this range that knowledge of inter-atomic forces is least adequate. Thus differences of opinion exist as to the extent to which spike effects and cascade effects are involved. The question will not be settled, on theoretical grounds at least, until more is known about the interaction energy of a pair of atoms at intermediate separations and until more refined methods are available for calculating the results of this interaction within a crystal lattice. Numerical procedures offer some hope, and will be discussed in section 3.21.

Ionizing radiation can also displace atoms and, again, several mechanisms have been proposed. In samples exposed to energetic gamma radiation, Compton and photoelectrons will be produced and will cause internal bombardment of sufficient vigor to displace atoms by direct knock-on processes (Dienes and Vineyard, 1957; Owen and Holmes, 1959). This mechanism has been experimentally

demonstrated with metals and seems to be reasonable theoretically. In a reactor, in which fast neutrons and gammas ordinarily are present with comparable fluxes, the displacements produced by gammas through this mechanism will be much less numerous than those produced by neutrons.

Two other mechanisms have been proposed to explain the production of displaced atoms in ionically bonded crystals by ionizing radiation. One mechanism, proposed by Seitz, involves the production of excitons which tend to be annihilated at jogs on dislocation lines. The energy released in annihilation may then cause a vacancy to be "boiled off" from the jog. This mechanism may be operative even with irradiation by ultraviolet light. A second mechanism, proposed by Varley, considers the consequences of multiple ionization. If a Cl^- ion in an NaCl lattice should lose two or more electrons under the action of ionizing radiation, it would find itself positively charged and occupying a lattice site which has a positive electrostatic potential. It thus will tend to be expelled to an interstitial position, where it may recapture an electron and migrate away. Both of these mechanisms may be operative, but the question of their relative efficiencies has not been settled, and more theoretical and experimental work is needed to resolve the matter. The question of the mobility of a neutral halogen atom in an interstitial position in an alkali halide lattice needs to be considered in the light of the evidence from magnetic resonance experiments that, at some temperatures, the stable configuration in KCl is that of the Cl_2 molecule, oriented in crowdion fashion along $\langle 110 \rangle$.

2.2. Models of Radiation Effects: Properties of Lattice Defects in Metals

In addition to the need for more precise models for the processes of production of lattice defects by radiation, better models are also required to describe the detailed behavior of the various lattice defects and the manner in which these defects alter the physical properties of materials. In metals, most physical properties are altered in a qualitatively similar manner by different types of lattice defects. In other solids, there is a similar ambiguity. Consequently, the lack of adequate defect-production models together with the inadequacy of our quantitative understanding of the behavior of defects have made difficult the interpretation of physical-property changes accompanying both production of and recovery from lattice damage. Controversy still exists as to the interpretation of much of the experimental data. In this and in the follow-

ing two sections, the current state of knowledge of the elemental lattice defects in solids produced by radiation is discussed.

Most of the work directed at gaining a basic understanding of radiation effects in metals has been done with the noble metals and some of their alloys. It has been the hope of researchers in this field that the series of metals, Cu, Ag, and Au, would exhibit similar behavior, with appropriate allowance for the effect of mass and atomic number on the damage rate. Early support for this arose from the electrical resistivity measurements of the damage rates ($d\rho/d\phi$) in these three metals, produced by low temperature 8 Mev deuteron irradiation (Cooper, et al., 1955). Recovery studies showed, however, that the fractional amount of recovery in each of the various recovery stages (numbered I through V) varied considerably from Cu to Au, with Ag being generally intermediate. Stage I recovery is largest in Cu and is almost nonexistent in Au, while Stage III exhibits the reverse trend. The sum of Stages I and III remains nearly constant, implying the existence of some relationship between them. It is too early, perhaps, to draw any conclusion as to the nature of this relationship. It indicates, however, that a study of the differences in the recovery behavior of Cu, Ag, and Au may be more informative than a study of a single metal.

2.2.1. Frenkel Defects

The two defects of most immediate concern in radiation damage are interstitial atoms and vacant lattice sites. The combined effect of an interstitial and a vacancy (Frenkel defect) on several physical properties is best known in copper. As pointed out by Seeger (1958), if the lattice strain around point defects is assumed to be that predicted by first order elasticity theory, the measurements of the group at the University of Illinois (Simmons and Balluffi, 1958; Vook and Wert, 1958; Magnuson, et al., 1958) on the recovery of the fractional change of lattice parameter $\Delta a/a$, length $\Delta l/l$, and resistivity $\Delta\rho$, can, within the limits of experimental uncertainty, be used to place limits on the resistivity of a Frenkel defect in Cu. These results indicate that the recovery occurring in copper in Stages I, II, and III, following deuterons irradiation, consists chiefly of interstitial-vacancy annihilation. Corbett, Smith and Walker (1959) conclude the same to be true following electron irradiation. The value for the resistivity of Frenkel defects is estimated to be about $2.9 \mu\Omega$ cm per atom percent Frenkel defects.

Corbett, et al., have shown that Stage I consists of at least five substages, including at least three distinct close interstitial-

vacancy pair recovery peaks (Stages I_A , I_B , and I_O) with different activation energies, and two recovery peaks (I_D and I_E) with identical activation energies. These last two substages have been attributed to correlated and uncorrelated interstitial-vacancy pair annihilation, respectively, resulting from the free migration of one of the defects, presumably the interstitial. While the close pairs exhibit different recovery kinetics than do separated Frenkel defects, the available evidence indicates that their static effects on most physical properties are not appreciably different.

The measurements of the release of stored energy in copper in Stage I following electron irradiation (Meehan and Sosin, 1959), support the interstitial-vacancy annihilation model and yield an energy of 4 to 5 eV per Frenkel defect. Reactor irradiations of copper at Oak Ridge (Blewitt, et al., 1956, 1959), however, have led to a much lower release of stored energy per unit of resistivity recovery in Stage I. This is a tantalizing discrepancy which seems to indicate that Stage I recovery after neutron irradiation occurs by a mechanism other than interstitial-vacancy annihilation.

2.2.2. Vacancies

The theoretical configuration of a vacancy in copper seems to be well established. When a single lattice site is vacant, the nearest neighbors relax inwardly in a symmetrical fashion a moderate amount; the second shell of neighbors relaxes outwardly very slightly, and the third inwardly. Beyond this shell the relaxation is slight and can be treated by conventional anisotropic elasticity theory.

The formation energy of a vacancy in copper has been calculated to be close to 1 eV (Tewordt, 1958). Whether it is slightly more or slightly less is still an unanswered question. For gold, it is almost exactly 1.0 eV (Bauerle and Koehler, 1957). Values have been determined for other metals as well (sec. 3.26). One reason for the lack of an experimental value for the vacancy formation energy in copper may derive from the unanswered question as to whether vacancy migration occurs in Stage III with an activation energy of 0.6–0.7 eV, or in Stage IV with an activation energy of 1.0–1.2 eV. Since the sum of the migration and formation energies should equal the activation energy of self-diffusion, 2.05 eV, the formation energy in copper should be either 0.9 eV or 1.4 eV. The former value implies somewhat higher thermodynamic equilibrium concentrations of vacancies at elevated temperatures in copper than in gold, while the latter indicates substantially lower concentrations. The migration energy of vacancies in gold is 0.82 eV. Regard-

less of which recovery stage in copper is chosen for vacancy migration, it is evident that close similarity does not exist between copper and gold in regard to the energies associated with vacancies.

The resistivity of vacancies in copper appears to be about $1.3\mu\Omega$ cm per atom percent vacancies. This value is supported by a number of theoretical calculations.

Bauerle and Koehler (1957) have demonstrated that vacancies in gold show a tendency to agglomerate, and that there exists at least one type of a multiple vacancy with an activation energy for motion of 0.60 eV or lower. There are theoretical indications that divacancies may be more mobile than single vacancies in copper. It is probable, therefore, that such defects contribute to Stage II recovery in copper and similar metals after irradiation with particles which are capable of producing multiple displacements. A recent important calculation (Damask, et al., 1959) reveals that the trivacancy, in contrast to the divacancy, is quite immobile in the close-packed lattice, but this result has not yet been fitted into the systematics of recovery processes.

2.2.3. Interstitials

The configuration of an interstitial atom and its migration mechanism in copper have received a considerable amount of theoretical attention. It is probable that one of the two configurations proposed by Huntington (1953) is the lowest energy configuration and that the interstitialcy migration mechanism should operate for the thermal movement of this configuration. There is still a possibility, however, that another configuration, the crowdion (Paneth, 1950; Tewordt, 1958), may exist as a metastable structure, even though its energy seems to be somewhat higher than that of the symmetric interstitial. Furthermore, it is possible that irradiation is an efficient means of producing crowdions. The relative production rates of crowdions and symmetric interstitials by irradiation, if both are produced, should be strongly dependent on type of irradiation, material irradiated and energy of irradiation, since the "focusing" of energy along close-packed directions depends strongly on both energy and atomic size (Sosin, et al., 1959). Even if a crowdion cannot exist as a metastable static configuration, the possibility must still be considered that it may exist transiently as a dynamic crowdion (sec. 2.1).

Corbett, et al. (1959), and Magnuson, et al. (1958), have shown that in copper an interstitial migrates with an activation energy of 0.12 eV. While they ascribe this energy to the migration of symmetric interstitials by the interstitialcy mechanism, others have

preferred to ascribe it to crowdion migration (Seeger, 1958; Sosin, 1959). The second group then ascribes the 0.6 ev activation energy of Stage III to the migration of symmetric interstitials, whereas the former ascribes this activation energy to vacancy migration. The arguments in support of each of these alternative assignments are given by the respective authors. Further work is being done at several laboratories in an effort to resolve this question for copper, and it seems probable that in the near future the correct assignment will be unambiguously demonstrated.

Little is known about the behavior of interstitials in other metals. Stage I recovery in gold, for example, appears to be either absent or much smaller than in copper. This may mean that the interstitial configuration which migrates with an activation energy of 0.12 ev in copper does not even exist in gold, or that it might exist and have a considerably higher or lower activation energy for migration. Stage I in silver has not been studied in sufficient detail to show whether or not free migration of a defect through the lattice can occur. It appears, therefore, that while the achievement of a satisfactory model for the behavior of interstitials in copper will certainly be helpful, it will probably not indicate immediately the solution to the problem with other metals, even for those most similar to copper.

2.2.4. Displacement Spikes

Homogenization of alloys apparently can be produced by any radiation capable of producing primary knock-on atoms with energies in the 10^4 ev range. Bleiberg, et al. (1956), and Konobeevsky (1956) have demonstrated that the homogenization produced by fission fragments in U-Mo and U-Nb alloys can result in the transformation of both these alloys to the metastable gamma phase. Both neutron irradiation and heavy-charged particle irradiation cause localized disordering of ordered alloys (the experiments are reviewed by Dienes and Vineyard, 1957). The displacement spike concept has been used to explain both of these phenomena, the primary effect of displacement spikes in each instance being the forced stirring of the atoms within small regions of about 10^5 to 10^6 Å³ in volume. Other mechanisms have also been proposed to account for these observations. Konobeevsky utilizes a thermal spike concept to explain the phase reversal in U—10 w/o Mo, and Kinchin and Pease (1955) apply the concept of replacement collisions to account for disordering of Ni₃Mn by neutron irradiation. Neither of these two concepts, however, is capable of explaining both phenomena, nor can replacement collisions account for the

disordering rates of Cu_3Au by charged particle irradiation (Brinkman, unpublished). The widely differing masses of copper and gold atoms make it unlikely that enough replacement collisions occur in Cu_3Au to account for the observed disordering rate resulting from proton irradiation, where the average knock-on has a relatively low energy. The displacement spike mechanism seems to be favored at present, therefore, as the mechanism responsible for these phenomena. It must still be recognized, however, that replacement collisions may contribute to disordering, particularly in alloys such as Ni_3Mn . In this instance, the masses of the two atoms are nearly equal, and energy transfer therefore is more favorable, so that a large number of replacement collisions can result from the production of a single knock-on atom.

2.2.5. Depleted Zones

Seeger (1958) has postulated that the production of "depleted zones" occurs in a metallic lattice as a result of high energy primary knock-ons (sec. 2.1). These zones are regions from which atoms have been carried long distances by means of a dynamic crowdion mechanism, leaving behind a deficiency of atoms. The material in such zones may have a different crystal structure from the surrounding lattice. Material of a structure which would be identified as representing a lower density phase has never been observed, although Gonser and Okkerse (1957) have obtained evidence for the formation of a *higher* density phase in GaSb and InSb (see also sec. 2.6). The depleted zone model does, however, offer an attractive explanation for radiation hardening in copper and for the fact that the hardening does not fully recover until temperatures of about 400°C are reached. It also provides a different mechanism for Stage I resistivity recovery after neutron damage as compared to recovery following electron irradiation, as well as an explanation for the low value of stored energy release observed in Stage I following neutron irradiation, relative to that found after electron irradiation (sec. 2.2.1).

Such depleted zones cannot explain disordering rates in ordered alloys, which displacement spikes are able to do, as pointed out earlier; however, displacement spikes, in themselves, cannot account for the observed radiation hardening. It is conceivable, therefore, that displacement spikes and depleted zones are two aspects of a more complex spike which disorders a region some $50\text{--}100\text{\AA}$ diameter and leaves a smaller zone of depleted material, some $10\text{--}30\text{\AA}$ in diameter, in its center. Certainly, considerably more work is needed in order to investigate the detailed nature

of the spikes produced by high energy primary knock-ons in metals.

2.3. Models of Radiation Effects: Properties of Lattice Defects in Non-Metals

2.3.1. Semiconductors

The most prominent feature of lattice defects introduced into semiconductors by irradiation is their action as donors or acceptors. A model proposed by James and Lark-Horovitz (1951) for the electronic energy levels of vacancies and interstitials in diamond-structure semiconductors has been often used for interpreting experimental results. In essence, this model is based upon the assumptions that an interstitial atom may be multiply ionized, i.e., act as a multiple donor, and that a vacancy may capture more than one extra electron, thereby acting as a multiple acceptor. When the interstitial and vacancy concentrations are equal, the role played by the various energy levels corresponding to the different possible charge states of the point defects will depend on the relative positions of the interstitial states with respect to the vacancy states. For example, it is generally assumed that the interstitial may lose as many as two electrons to the lower levels of the vacancy, thereby becoming doubly ionized. It then assumes the role of the acceptor center, and the vacancy becomes the donor. Although it is possible to use the qualitative features of this model to explain experimentally observed defect states, it has a number of deficiencies: (1) It is not unique, since other energy level structures can be proposed which are equally satisfactory. (2) Being primarily a phenomenological model, it is not amenable to quantitative computations. (3) Finally, it is strictly applicable only to isolated point defects and requires extensive revision for application to composite damage sites. More recently, Blount (1959) has proposed a model in which both interstitials and vacancies may act as amphoteric centers, i.e., both may function as either acceptors or donors, depending on the position of the Fermi level. Using arguments based on the properties of sp^3 orbitals, Blount found that a reasonable positioning of these states in the forbidden energy gap corresponds to certain observed radiation-defect states. Once again, a quantitative calculation is exceedingly difficult and has not been carried out.

The only rigorous attempt to calculate the electronic energy levels of a point defect in a valence crystal was that of Coulson and Kearsley (1957), who applied group theoretical methods to a

vacancy in diamond. They found that the first optically excited state of a neutral vacancy lies about 2 eV above the ground state, and that the neutral vacancy is apparently more stable in the diamond structure than a vacancy which has captured an electron. An extension of such computations to interstitials in diamond, as well as to interstitials and vacancies in silicon and germanium, would be exceedingly valuable.

In addition to the energy levels associated with isolated point defects, some consideration has been given to the effects of high local concentrations of defects which result from fission neutron bombardment. Gossick and Crawford (1957, 1959) have considered the nature of the electrostatic field surrounding regions (~ 100 to 200\AA in radius) containing high defect density ($\sim 10^{19}\text{ cm}^{-3}$) in n-type germanium or n- or p- type silicon. It was found that in order to sustain a converted region (p-type in n-type Ge) of this size, a positive space charge must extend an appreciable distance into the n-type matrix. Depending on the density of defects in the disordered region associated with a displacement cascade and the density of donors in the n-type Ge matrix, the region of positive space charge and its associated potential gradient may extend more than an order of magnitude beyond the disordered region. Because of the large values of electrostatic potential in most of this region, it is denuded of conduction electrons and acts as insulator. This model has been used with some success to explain distributions in positions of defect states as well as the apparently anomalous behavior of mobility in neutron-irradiated n-type germanium (Crawford and Cleland, 1959). Application of this model to expected intrinsic regions in neutron-bombarded Si also seems to show some promise.

2.3.2. Insulators

Apart from these defect-level models for germanium and the various models proposed for defects in the alkali halides (Seitz, 1954), little systematic effort has been spent in devising specific theoretical models for radiation-induced defects in other non-metallic systems. Most interpretations, when available, are in terms of specific considerations based on the properties and physical behavior of the material in question. The work of Coulson and Kearsley on diamond, as mentioned above, is a notable exception. When reduced to its essentials, the problems of energy-level models for radiation-induced lattice damage in non-metals are not appreciably different from the general problems of determining the wave functions and calculating the matrix elements for transitions be-

tween states for the various stationary states of imperfections in non-metals. This problem will continue to be a puzzling one for some time to come. These remarks serve to indicate the large gaps in our theoretical understanding of the nature of defects in non-metallic systems.

2.4. Models of Radiation Effects: Production Processes and Properties of Lattice Defects in Materials Suffering Transmutations

Materials which suffer prominent nuclear reactions can be considered in a separate category, including the fissionable elements, e.g., U, Th, Pu, and those materials which react with the bombarding particle, emitting elementary particles and undergoing transmutation, e.g., B, Li, Be. This latter comprises, of course, a large class of materials. In addition to the normal atomic displacements, displacements are produced by the emitted particle with the introduction of impurity atoms, but when the transmutations are less numerous than the natural impurities only the displacement effects are important. The effects are thus relatively more important in pure materials, e.g., semiconductors. The damage resulting from atomic displacement saturates at high doses, but since the number of transmutations is proportional to the dose, the effects preponderate at higher dosage. As higher particle doses become available, the effects will become increasingly important.

Although the nuclear physics and the cross sections of the reactions which occur when materials are bombarded with atomic particles are known, the effect of the reactions upon their surroundings can only be outlined. More precise models are necessary before the effects can be adequately described.

These transmutable materials undergo all the atomic displacements experienced by other non-reactive materials, i.e., vacancies, interstitials, displacement spikes, and depleted zones may be produced (secs. 2.1 to 2.3). The large energy release which often accompanies the transmutation (each fission fragment in uranium possesses $\sim 10^8$ ev) produces damage similar to, but more extensive than, that due to primary knock-ons produced by artificially accelerated particles. For these latter particles, it appears that a spike model is the most applicable, although this has taken various forms, e.g., thermal spike, displacement spike, and depleted zone (secs. 2.1 and 2.2).

This intense damage produces growth of anisotropic alpha-uranium crystals in the [010] direction with an equivalent contraction in the [100] direction. Several attempts have been made

to explain this behavior involving either spike theories or point-defect theories. Neither of these is very satisfactory, for neither offers an easy explanation of the lack of crystal alteration in the third dimension. The spike theories also do not account for the observations that the growth is structure sensitive (e.g., cold-worked crystals grow faster) and that the growth rate decreases at low temperatures. The other theories must assume anisotropic behavior of point defects, with either the vacancies and interstitials migrating along mutually perpendicular axes before annihilating at boundaries, or annihilating upon dislocations with mutually perpendicular Burgers vectors. Also the number of point defects necessary for the mass transfer is large ($\sim 10^8$ per fission) so that the anisotropic process must be an efficient one. Since the two theories together can explain most of the observations qualitatively, a theory integrating both concepts may be more successful.

The transmutation products are in themselves additional lattice defects and because their effects are dictated mainly by their chemical nature and that of the matrix, each situation must be considered individually. However, in many materials the effects are negligible because the cross sections are so small that the impurity build-up is less than the natural level of impurities present. For example, in an integrated neutron flux of $10^{22}/\text{cm}^2$ (a very high exposure) the cross section must be greater than 10^{-2} barns in order for the induced impurity level to exceed a concentration of 10^{-4} .

The extra atoms that are formed produce a slight dilatation which cannot be annealed out. The effect of the extra atoms becomes marked when they are highly insoluble, as embrittlement may then occur. The inert gas atoms are a very important class of products, as they are both very insoluble and tend to precipitate in a gaseous form, producing swelling. The factors involved in this swelling can be listed as:

1. Formation of inert gas atoms
2. Solubility of these atoms
3. Migration of the atoms
4. Nucleation of bubbles
5. Stability of the bubbles
6. Creep strength of the matrix

The first factor involves straight-forward nuclear physics and is well understood. Calculations of energies of solution of various inert gas atoms in copper indicate that the larger atoms Xe and Kr would go into substitutional solution by the capture of one or more vacancies (Rimmer and Cottrell, 1957). Several measure-

ments of the diffusion rates of the inert gas atoms have been made, but the effect of radiation environment upon diffusion rates is not known. Experiments with copper bombarded with alpha-particles from a cyclotron (Barnes, et al., 1958) have demonstrated that the helium atoms so injected diffuse in a substitutional manner. These latter experiments also have shown that the precipitation of these atoms as gas bubbles is governed by the available supply of vacancies. At temperatures for which the supply of vacancies is inadequate, volume increases are restricted (Barnes and Greenwood, 1958). The nucleation of the bubbles is thought to be on a fine scale ($\sim 10^{13}$ bubbles cm^{-3}) and to occur on dislocation lines existing in the material. The stability of the bubbles also is important. In the initial state they are so small that surface energy affects their volume. It appears that under certain conditions helium is able to redissolve, leading to coarsening of some of the bubbles and large swelling. It is important to establish the stability of other gases, particularly in the radiation environment.

2.5. Existing Experimental Coverage

The next four sections attempt to provide a picture of the present state of experimental investigations in the field and an indication of the areas that have not yet been explored. From among the great number of experiments that have been performed we have selected those which we regard as fundamental for testing theoretical models of radiation effects, i.e., those in which conditions have been sufficiently simple and well controlled so that it might be possible to achieve full understanding. Two striking facts emerge from this compilation. In no instance is it possible to state that complete and unarguable understanding of a radiation damage phenomenon has been achieved—even the best understood cases are somewhat hypothetical. Secondly, the materials that have been carefully investigated are exceedingly few.

2.5.1. Metals and Alloys

Experiments on metals and alloys have been summarized in the following Tables, with pure metals covered in Table 1 and alloys in Table 2. Table 1 is subdivided into sections entitled "Explanatory Research" (covering the more qualitative work) and "Quantitative Measurements." The work summarized in Table 2 is qualitative rather than quantitative and is organized by sections covering "Dilute Alloys," "Order-Disorder Systems," and "Solid Solutions."

The first column in each table cites the metal and the reference to the detailed report of the experiment; (S) indicates that the specimens were single crystals. In the next column the purity is designated by quoting the residual resistivity, even though the impurity content may be actually greater than indicated by the

TABLE 1
EXPERIMENTAL RESULTS ON PURE METALS
A. EXPLORATORY RESEARCH

Metal and reference	Purity (ohm cm)	Rad.	Integ. flux	Rad. T	Property measured	Type of anneal	Results
Al (S) (1)	1×10^{-8}	n	$4 \times 10^{17}f$	19.2°K	Energy	Cont.Heat	$\Delta E < 0.1$ cal/mole
Al (S) (1)	1.1×10^{-8}	n	$4 \times 10^{17}f$	18°K	Resist.	3 Min pulse (M 14°K)	begins 30°K, 1/2 below 50°K
Fe (1)	3.1×10^{-8}	n	$4 \times 10^{17}f$	20°K	Resist.	3 Min pulse (M 14°K)	begins 75°K, 4/10 below 120°K
Co (1)	3.1×10^{-7}	n	$4 \times 10^{17}f$	18°K	Resist.	3 Min pulse (M 14°K)	
Ni (1)	1.9×10^{-8}	n	$4 \times 10^{17}f$	18°K	Resist.	3 Min pulse (M 14°K)	begins 63°K
Cu (2)	1×10^{-8}	10 Mev d	1.0×10^{17}	15°K	Resist.	Cont. Heat	1/2 below 45°K 1/4 garbage
Zn (1)	2.4×10^{-8}	n	$4 \times 10^{17}f$	19°K	Resist.	3 Min pulse (M 14°K)	begins 80°K lots garbage
Ag (2)	1×10^{-8}	10 Mev d	1.0×10^{17}	14°K	Resist.	Cont. Heat	1/4 below 35°K 3/10 garbage
Sn (3)		n	$1 \times 10^{18}f$	80°K	Volume	Hold at -50.3°K	defects nucleate phase change
Pt (1)	1.5×10^{-7}	n	$4 \times 10^{17}f$	15°K	Resist.	3 Min pulse	3/10 below 65°K
Au (2)	1×10^{-8}	10 Mev d	1.0×10^{17}	14°K	Resist.	Cont. Heat	No low T drop, 1/2 garbage
Au (4)		n		330°K	Microstructure	None	observed spots 50-100 Å diam. in electron microscope
Au (5)		p	3.4×10^{18}	300°K	Atoms Ejected	None	Au atoms ejected in close packed directions

¹ Blewitt, T. H., Coltman, R. R., Holmes, D. K., Noggle, T. S. 1957, Creep and Recovery (Am. Soc. for Metals) p. 84.

² Cooper, G., Koehler, J. and Marx J., 1955, Phys. Rev. 97, 599.

³ Fleeman, J., Dienes, G. J., 1955, J. Appl. Phys. 26, 652.

⁴ Noggle, T. S., 1959, Bull. Am. Phys. Soc. 4, No. 3, 137.

⁵ Thompson, M. W., 1959, Phil. Mag. 4, 189.

TABLE 1—Continued
EXPERIMENTAL RESULTS ON PURE METALS
B. QUANTITATIVE MEASUREMENTS

Metal and reference	Purity (ohm cm)	Rad.	Integ. flux	Rad. T	Property measured	Type of anneal	Results
Cu (1)	1.5×10^{-9}	1.37 Mev e	8.4×10^{16}	13°K	Resist.	10 Min pulse (M 4.2°K)	$E = 22 \pm 3$ ev 9/10 below 65°K
Cu (2)	1.5×10^{-9}	1.4 Mev e	3.3×10^{16}	12.0°K	Resist.	10 Min pulse (M 4.2°K)	3 close pair processes (.05, .085, .095 ev); 1 interstit. mig. ($E_M = .12$ ev)
Cu (3)	1.5×10^{-9}	1.4 Mev e	8×10^{17}	80°K (then 3.3×10^{16} at 20°K)	Resist.	10 Min pulse (M 4.2°K)	get long dist. migration at 55°K
Cu (4)	1.4×10^{-9}	10.7 Mev d	2.5×10^{16}	12°K	Resist.	Isotherm (1.5°K steps)	3 close pair processes (.048, .080, .091 ev) 1 interstit. mig. ($E_M = .113$ ev)
Cu (5) (8)	5.4×10^{-9}	7 Mev d	6.3×10^{16}	10°K	Lattice Param.	Cont. Heat	No line broad. $\Delta d/d = 4.1 \times 10^{-21}$ /d/cm ² ; annealing is like that of ρ and $\Delta v/v$.
Cu (6)	1.5×10^{-9}	8.5 Mev d	7.7×10^{16}	17°K	Length	Cont. Heat	$\Delta v/v = 3.8 \times 10^{-21}/d/cm^2$
Cu (7)	1.5×10^{-9}	n	4×10^{17} f	18°K	Energy Release	Cont. Heat	$\Delta E/\Delta \rho = 1.7$ cal/gm/ $\mu\Omega$ cm below 60°K
Cu (8)	1.5×10^{-9}	n	5×10^{17} f	3.9°K	Resist.	30 sec pulse (M 3.9°K)	Fast annealing, 1/2 in 1 MIN.
Cu (9)		Cu		20°C	Count Recoils		Range of recoils in Cu is 180Å. Bohr potential too small at intermediate distances.
Cu (8) (10)	1.5×10^{-9}	n	4×10^{17} f	21°K	Young's Mod. and Internal Friction		Get slowly increasing pinning of dislocations at 21°K
Cu (11)	1.5×10^{-9}	10 Mev d	1.0×10^{17}	95°K	Resist.	Isotherm (12°K steps)	E_M increases (93°K to 210°K) $E_M = 0.68$ ev (230°K to 310°K)

TABLE 1—Continued
EXPERIMENTAL RESULTS ON PURE METALS—Continued
B. QUANTITATIVE MEASUREMENTS—Continued

Metal and reference	Purity (ohm cm)	Rad.	Integ. flux	Rad. T	Property measured	Type of anneal	Results
Ag (4)	1.85×10^{-9}	10.7 Mev d	2.5×10^{16}	12°K	Resist.	Isotherm (12°K steps)	2 close pair processes (.045, .058 ev); 1 interstit. mig. ($E_M = 0.079$ ev)
Cu (12)	3.4×10^{-9}	1.25 Mev e	3.4×10^{18}	80°K	Resist.	Isotherm 298°K 5 Min pulse (15°K steps)	$E_M = 0.60$ ev (275°K to 325°K)
Cu (13)	$\sim 3.4 \times 10^{-9}$	1.2 Mev e	9×10^{17}	20°K	Energy Release	Cont. Heat	$\Delta E / \Delta \rho \cong 5.4$ cal/gm/ $\mu\Omega$ cm below 50°K

1. Corbett, J. W., Denney, J. M., Fiske, M. D., and Walker, R. M., 1957, Phys. Rev. 108, 954.
2. Corbett, J. W., Smith, R. B., and Walker, R. M., Phys. Rev. (to appear).
3. Corbett, J. W., Smith, R. B., and Walker, R. M., Phys. Rev. (to appear).
4. Magnuson, G. L., Palmer, W., and Koehler, J. S., 1958, Phys. Rev. 109, 1990.
5. Simmons, R. O. and Balluffi, R. W., 1958, Phys. Rev. 109, 1142.
6. Vook, R. and Wert, C., 1958, Phys. Rev. 109, 1142.
7. Blewitt, T. H., 1958, Symposium on Vacancies and Point Defects, Harwell.
8. Blewitt, T. H., Coltman, R. R., Klabunde, C. E., Redman, J. K., Diehl, J., 1959, Bull. Am. Phys. Soc. 4, No. 3.
9. Schmitt and Sharp, 1958, Phys. Rev. Lett. 1, 445.
10. Thompson, D. O. Blewitt, T. H., and Holmes, D. K., 1957, J. Appl. Phys. 28, 742.
11. Overhauser, A., 1953, Phys. Rev. 91, 448.
12. Meechan, C. J. and Brinkman, J. A., 1956, Phys. Rev. 103, 1193.
13. Meechan, C. J. and Sosin, A., 1959, Phys. Rev. 113, 422.

residual resistivity. Next are, in order, the type of irradiation (n means pile irradiation) and the integrated flux (f means fast neutrons) in particles per cm². The temperature of irradiation is specified under "Rad. T." The property measured, and the type of anneal are listed next (Cont. heat means continuous heating; 3 MIN. PULSE means pulsing from the measuring temperature, holding 3 minutes at a given annealing temperature, cooling quickly, and measuring at the specified temperature); the measuring temperature (M 14°K) is given in the same column. The last column reports the results, e.g., ΔE is the energy released on warming to 60°K; qualitative remarks concerning the annealing (begin 30°K, 1/2 below 50°K); E_d represents the minimum energy which must be given to a lattice atom to displace it; E_M means activation energy required for migration; $\Delta d/d$ is the fractional change in lattice

TABLE 2
EXPERIMENTAL RESULTS FOR ALLOYS
A. DILUTE ALLOYS

Metal and reference	Imp. Conc.	Rad.	Integ. flux	Rad. T.	Property measured	Type of anneal	Results
Cu (1)	0.1% Be	n	4×10^{17}	14.8°K	Resist.	3 Min pulse (M 10°K)	Alloying prevents 2/3 of 40°K ann.
Cu (1)	0.1% Si	n	4×10^{17}	14.5°K	Resist.	3 Min pulse (M 10°K)	Alloying prevents 3/4 of 40°K ann.
Cu (1)	1.0% Ni and 0.1% Ni	n	4×10^{17}	14.5°K	Resist.	3 Min pulse (M 10°K)	No effect
Cu (1)	0.17% Zn	n	4×10^{17}	14.5°K	Resist.	3 Min pulse (M 10°K)	No effect
Cu (1)	0.1% Au	n	4×10^{17}	14.5°K	Resist.	3 Min pulse (M 10°K)	Alloying prevents 1/2 of 40°K ann.

B. ORDER-DISORDER SYSTEMS

Cu ₃ Mn (2)	Mostly ordered (Initial $\rho = 2.5151 \times 10^{-6} \Omega \text{ cm}$)	n	4×10^{17}	18°K	Resist.		$\Delta \rho_{irr} = 4.8 \times 10^{-6} \Omega \text{ cm}$, slight low T ann.
Cu ₃ Mn (2)	Mostly disordered (Initial $\rho = 4.7196 \times 10^{-6} \Omega \text{ cm}$)	n	4×10^{17}	18°K	Resist.		$\Delta \rho_{irr} = 4.6 \times 10^{-6} \Omega \text{ cm}$, slight low T ann.
Cu ₃ Mn (2)	Partly ordered	n		110°K	Resist.	Pulse (M 77.3°K)	Rapid ordering above 150°K
Cu ₃ Mn (3)	Mostly disordered (Initial $\rho = 7.92 \times 10^{-6} \Omega \text{ cm}$)	n	1.4×10^{17}	150°C	Resist.		Saw ordering during irradiation.
Cu ₃ Mn (4)	Partly ordered	1 Mev e	3×10^{17}	43°C	Resist.	Isotherm 100–130°C	Got drop in ρ on ann.
Cu ₃ Mn (5)	Partly ordered	Co γ	6×10^{16}	0°C	Resist.	Isotherm 128.1°C	Got drop in ρ on ann.
μ Brass (6)	Ordered	33 Mev d		173°K	Resist.	Pulse Heat	Can disorder by irradiation.
Ni ₃ Mn (7)	Ordered	n	5×10^{19}	50°C	Resist. and Mag.	Induct.	Anomalously rapid disordering.

TABLE 2—Continued
EXPERIMENTAL RESULTS FOR ALLOYS—Continued
C. SOLID SOLUTIONS

Metal and reference	Imp. Conc.	Rad.	Integ. flux	Rad. T.	Property measured	Type of anneal	Results
α Brass (8)	30% Zn	2 Mev e	2.6×10^{16}	50°C	Resist.		ρ drops; rate of short range ordering greatly enhanced during irradiation.
α Brass (9)	30% Zn	n	10^{16}	0°C-300°C	Resist.		
CuAl (10)	6% to 15% Al	n	3.5×10^{16}	35°C-120°C	Resist.	Hold at 35°C	ρ drops (large in 15% alloy)

1. Blewitt, T. H., Coltman, R. R., Klabunde, C. E., Noggle, T. S., 1957, *J. Appl. Phys.* **28**, 639.
2. Blewitt, T. H., Coltman, R. R., Holmes, D. K., Noggle, T. S., 1957 *Creep and Recovery* (Am. Soc. for Metals), 106-107.
3. Blewitt, T. H., and Coltman, R. R., 1954, *Acta Met.* **2**, 549.
4. Adam, J., Green, A., and Dugdale, R. A., 1952, *Phil. Mag.* **45**, 1216.
5. Dugdale, R. A., 1955, *Report of the Bristol Conference on Defects in Crystalline Solids* (Physical Society, London), 246.
6. Eggleston, R. R. and Bowman, F. E., *J. Appl. Phys.* **24**, 299.
7. Aronin, L. R., 1954, *J. Appl. Phys.* **25**, 344.
8. Damask, A. C., 1958, *J. Phys. Chem. Solids* **4**, 177.
9. Dienes, G. J. and Damask, A. C., 1958, *J. Appl. Phys.* **29**, 1713.
10. Wechsler, M. S. and Kernohan, R. H., 1958, *J. Phys. Chem. Solids* **7**, 307.

parameter; $\Delta v/v$ is the fractional volume change; ρ represents resistivity; "garbage" refers to the extended low temperature annealing in which definite stages are not discernible.

The experimental coverage still is far from complete. It is most comprehensive for the noble metals, especially for copper. However, even with copper reliable annealing data above room temperature on 99.999% pure material does not exist. In the other noble metals, various gaps in the experimental coverage exist; for example, no good isothermal annealing studies are available for gold, for any kind of irradiation over any range of temperature. The noble metals have been chosen for concentrated study because of the available theoretical and related experimental information concerning them.

High quality experimental work on other metals is very meager. As a result, the influence of crystal structure on the details of production and annealing is uncertain. Research on alloys also has been limited. Yet even the little that has been done has yielded rather suggestive results.

TABLE 3
SUMMARY OF STUDIES OF ELECTRONIC EFFECTS OF IRRADIATION
IN SEMICONDUCTORS

Material	Reference (See below)	Radiation	Temperature	Property investigated (a)
Diamond	(1)	2 Mev electron	80°K	R, α
Silicon	(2, 3)	Reactor neutrons	320°K	σ , R, μ , σ_p
Do	(4)	Reactor neutrons	320°K	α
Do	(5)	0.7 Mev electrons	300°K	τ , R, σ , μ
Do	(6)	Reactor neutrons	300°K	τ , R, σ , μ
Do	(7)	Co ⁶⁰ gamma rays	300°K	R, σ , μ
Do	(8)	4.5 Mev electron	300°K	R, σ , μ , E.S.R.
Do	(9)	Reactor neutrons	300°K	χ , R, σ
Germanium	(10, 11, 12)	Reactor neutrons	110°K to 320°K	R, σ , μ
Do	(13, 14)	Reactor neutrons	300°K	τ , R, σ
Do	(14, 15, 16)	Co ⁶⁰ gamma rays	310°K	τ , R, σ , μ
Do	(17)	Reactor neutrons	16°K	σ (Pulse anneal)
Do	(18)	Po ²¹⁰ alpha particles	10°K	σ , R, μ (Pulse anneal)
Do	(19)	1.1 and 0.35 Mev electrons	10°K	σ , R, μ (Pulse anneal)
Do	(20)	Reactor neutrons	300°K	χ , R, σ
Do	(21)	1.7 to 3.8 Mev neutrons	77°K	σ
Do	(22)	14 Mev neutrons	77°K, 300°K	τ , σ , R
Do	(23)	10 Mev deuterons	77°K, 300°K	σ , R, μ , σ_p
InSb	(24)	Reactor neutrons	150°K, 300°K	σ , R, μ
InSb	(8, 25)	Fast electrons	80°K, 200°K	σ , R, μ
InSb	(26)	Co ⁶⁰ gamma rays	77°K	σ
GaSb	(27)	Reactor neutrons	150°K, 300°K	σ , R, μ
InAs	(28)	Reactor neutrons	300°K	σ , R, μ
AlSb	(29)	Reactor neutrons	300°K	σ , R, μ , ρ_m
CdTe	(30)	Reactor neutrons	300°K	σ , R, μ

(a) Notation:

σ —	Electrical conductivity
R—	Hall coefficient
μ —	Mobility
τ —	Minority carrier lifetime
α —	Optical absorption
σ_p —	Photoconductivity
	(Spectral dependence)
χ —	Magnetic susceptibility
ρ_m —	Magneto resistance
E.S.R.—	Electron Spin Resonance

1. Mitchell, E. W. J., 1959, *J. Phys. Chem. Solids* **8**, 444.
2. Longo, T. A., 1957, Thesis (Dept. of Physics, Purdue University); Longo, T. A. and Lark-Horowitz, K., 1957, *Bull. Am. Phys. Soc.*, Series 2, 156.
3. Crawford, J. H. Jr., and Cleland, J. W., 1957, *Progress in Semiconductors*, Heywood and Co., London, Vol. 2, p. 67.
4. Fan, H. Y., and Ramdas, A. K., 1959, *J. Phys. Chem Solids*, **8**, 272.
5. Wertheim, G. K., 1957, *Phys. Rev.* **105**, 1780; 1958, *Phys. Rev.* **110**, 1272.
6. Wertheim, G. K., 1958, *Phys. Rev.* **111**, 1500.
7. Sonder, E. and Templeton, L. C., 1958, *Bull. Am. Phys. Soc.* **3**, 375.
8. Fan, H. Y., and Lark-Horowitz, K., 1958, *Semiconductors and Phosphors*, (Interscience Publishers, New York), p. 113.
9. Sonder, E. and Stevens, D.K., 1958, *Phys. Rev.* **110**, 1027.
10. Cleland, J. W., et al, 1951, *Phys. Rev.* **84**, 861.
11. Cleland, J. W., Crawford, J. H. Jr., and Pigg, J. C., 1955, *Phys. Rev.* **98**, 1742.
12. Cleland, J. W., Crawford, J. H. Jr., and Pigg, J. C., 1955, *Phys. Rev.* **99**, 1170.
13. Curtie, O. L., et al, 1957, *J. Appl. Phys.* **28**, 1161.
14. Curtis, O. L., Cleland, J. W., and Crawford, J. H. Jr., 1958, *J. Appl. Phys.* **29**, 1722.
15. Cleland, J. W., Crawford, J. H. Jr. and Holmes, D., 1956, *Phys. Rev.* **103**, 722.

16. Crawford, J. H. Jr., June 1958, Proc. of Int. Conf. on Solid State Physics, Brussels, (to be published).
17. Cleland, J. W. and Crawford, J. H. Jr., 1958, J. Appl. Phys. *29*, 149.
18. Gobeil, G. W., 1958, Phys. Rev. *112*, 722.
19. Mackay, J. W., Klontz, E. E., and Gobeil, G. W., 1959, Phys. Rev. Letters, *3*, 146.
20. Stevens, D. K., et al., 1955, Phys. Rev. *100*, 1084.
21. Ruby, S. L., Schup, F. O., and Wolley, E. D., 1958, Phys. Rev. *111*, 1493.
22. Curtis, O. L. and Cleland, J. W., J. Appl. Phys. (to be published).
23. Fan, H. Y. and Lark-Horowitz, K., 1958, *Effects of Radiation of Materials* (Reinhold, New York), pp. 159-190 (Edited by Harwood, et al.).
24. Cleland, J. W. and Crawford, J. H. Jr., 1955, Phys. Rev. *95*, 1177.
25. Eisen, F., July 1958, WADC Report 58-428.
26. Whitsett, C. A., 1958, Bull. Am. Phys. Soc. *3*, 142.
27. Cleland, J. W. and Crawford, J. H. Jr., 1955, Phys. Rev. *100*, 1614.
28. Cleland, J. W. and Crawford, J. H. Jr., 1958, Bull. Am. Phys. Soc. *3*, 142.
29. Reid, Beer, and Willardson, 1957, Bull. Am. Phys. Soc. *2*, 356.
30. Reid, Moody, and Willardson, 1957, Radiation Effects Information Center, Battelle Memorial Institute, REIC-1.
31. Hill, D. E. and Lark-Horowitz, K., 1958, Bull. Am. Phys. Soc. *3*, 142.

2.5.2. Semiconductors

Ever since early experiments suggested that there might be a direct correlation between defect-acceptor states and the Frenkel pairs produced by bombardment in germanium, diamond structure semiconductors, notably silicon and germanium, have received considerable attention in radiation effects investigations. Subsequent developments have shown that the early view, namely that the change in carrier concentration induced by a given exposure is a direct measure of the density of displaced atoms, is unduly optimistic. Nevertheless it is true that, because of the intimate relationship of electronic properties to lattice imperfections, semiconductors are excellent materials in which to probe the details of lattice damage produced by energetic particle bombardment. Not only are the concentrations of charge carriers altered by the localized states associated with isolated and composite defects, but the scattering of carriers by these charged imperfections affects all properties involving charge transport.

Those materials of the diamond-and-zinc-blende structure which have been investigated, together with the properties which have been used as indices of lattice damage, are listed in Table 3. Effects on the Fermi level for the various materials are listed in Table 4, and the positions of radiation-induced energy levels which have been experimentally detected are listed in Tables 5 and 6 for Ge and Si, respectively.

Studies on annealing of defect structures, particularly in germanium, also have been reported. Below 200°K relaxations in the electronic system, i.e., release of minority carriers from traps and their subsequent recombination with excess majority carriers, also produce large changes in the electrical property being meas-

TABLE 4
ELECTRON EFFECTS OF ENERGETIC RADIATIONS ON CERTAIN SEMICONDUCTORS

Material	Radiation used	Predominant role of defects	Effect on n-type	Effect on p-type
Ge	Fast neutrons, alpha particles, 10 Mev deuterons, electrons (< .3 Mev).	Acceptor	Converts to p-type	All p-type approach an intermediate limiting hole concentration.
	Gamma rays	Acceptor	Converts to p-type	High-conductivity p-type not appreciably affected.
Si	Fast neutrons, 10 Mev deuterons, electrons (< .3 Mev).	Acceptors and donors nearly equally effective.	Tends to intrinsic	Tends to intrinsic.
InSb	Fast neutrons	Donor	Approaches low limiting electron concentrations.	Converts to n-type.
	4.5 Mev electrons at 200°K and gamma rays at 77°K.	Acceptor	Converts to p-type	Small decrease in hole concentration.
GaSb	Fast neutrons	Acceptor	Electron concentration increases.	Approaches an intermediate limiting hole concentration.
InAs	Fast neutrons	Donor	Electron concentration increases.	Converts to n-type.
AlSb	Fast neutrons	Donor	Approaches limiting electron concentration.	Converts to n-type.
CdTe	Fast neutrons	Uncertain — large donor activity indicated.	Unknown	Large increase in resistivity.

ured. Therefore, interpretation of relaxations occurring during warming after low-temperature bombardment requires careful separation of annealing of defects from the electronic effects. Under certain circumstances, for example, where degenerate specimens are annealed in the absence of ionizing radiation, experimental study has yielded good information on pulse annealing peaks in germanium after 1.1 Mev electron (MacKay, et al., 1959) and after 4.5 Mev alpha-particle (Gobeli, 1958) bombardment near 10°K. A first-order relaxation was found at 35°K (0.04 ev activation energy) and a more complex peak was observed at 65°K (0.06 to 0.09 ev). Exposure to electrons of energy below or near the displacement threshold (< 0.6 Mev) at 10°K apparently accomplishes the same purpose as thermal annealing, since such exposure caused an appreciable recovery in both resistivity and Hall coefficient,

TABLE 5
BOMBARDMENT-INDUCED ENERGY LEVELS IN GERMANIUM

Radiation used	Energy level (ev) (+)	References (See Table 3)
Reactor neutrons	$E_C - 0.2$	(11, 13, 14)
	$E_V + 0.18$	(12)
	$E_V + 0.07$	(12)
14 Mev neutrons	$E_V + 0.32$	(22)
	$E_V + 0.13(b)$	(22)
10 Mev deuterons	$E_C - 0.02$	(23)
	$E_C - 0.10$	(23)
	$E_C - 0.23$	(23)
	$E_C - 0.35$	(23)
	$E_V - 0.2$	(23)
	$E_V - 0.08$	(23)
	$E_V - 0.02$ (Two Levels)	(23)
4.5 Mev electrons	$E_C - 0.1$	(23)
	$E_C - 0.24$	(23)
	$E_C - 0.35$ (c)	(23)
	$E_V + 0.22$	(23)
	$E_V + 0.1$	(23)
	Two states very near E_V	(23)
Co ⁶⁰ gamma rays	$E_C - 0.20$	(15)
	$E_V + 0.26$	(15)

TABLE 6
BOMBARDMENT-INDUCED ENERGY LEVELS IN SILICON

Radiation	Energy level (ev)	References (See Table 3)
Reactor neutrons	$E_V + 0.3$	(12)
	$E_V + 0.25$	(12)
	$E_V + 0.20$	(12)
	$E_V + 0.17$	(12)
10 Mev deuterons	$E_C - 0.025$	(2)
	$E_V - 0.065$	(2)
0.7 Mev electrons	$E_C - 0.16$	(5)
	$E_V + 0.29$	(5)
	Levels near E_C and E_V	(5)
4.5 Mev electrons	$E_C - 0.17$	(31)
	$E_C - 0.36$	(31)
	Levels near E_C and E_V	(31)

and little thermal recovery was observed subsequently. Annealing studies also have been made in the range above room temperature. From isothermal annealing behavior, the characteristic activation energy, presumably associated with the motion of interstitial atoms, was found to be 1.7 and 1.8 eV for electron bombarded Ge (Brown, et al., 1954) and Si (Bemski and Augustyniak, 1957), respectively, for temperatures as high as 400°C. Chang (1957) observed an activation energy of 1.7 eV for the annealing of the effect of neutron bombardment on the etching behavior of Ge.

Because of the extreme sensitivity of certain electronic properties of high-purity valence semiconductors, they are excellent materials in which to investigate very small concentrations of lattice damage. Though early studies (Klontz and Lark-Horovitz, 1952) of displacement energies indicate a value of $E_d = 31$ eV in Ge, Rappaport and Loferski (1955) have shown that permanent effects are produced, with small probability, for bombarding electron energies corresponding to E_d as low as 13 eV in both Si and Ge. Brown (1959) has investigated the effect of crystal orientation with respect to the direction of the primary electron beam on the yield of defects for electron energies near the displacement threshold (0.2 to 0.6 MeV). Also important are experiments involving 14 MeV neutrons from the $H^3(H^2,n)He^4$ reaction (Curtis and Cleland, to be published) and neutrons in the range 1.5 to 4.8 MeV from the $H^2(H^2,n)He^3$ reaction (Ruby, et al., 1958). In general, neutrons of these energies are obtained from ion accelerators and the available fluxes are so low that the sensitivity of minority carrier lifetime, conductivity, and Hall coefficient to small concentrations of defects make semiconductors invaluable for such investigations.

Since the spatial distribution of defects, i.e., fluctuations in local concentration, has important consequences for carrier scattering, semiconductors are also excellent media in which to investigate the detailed nature of defect configurations resulting from exposure to different energetic particles. Wertheim (1957, 1958) has combined investigations of mobility and lifetime in electron and neutron bombarded Si in exploration of the correlations in distance between defects. Furthermore, comparison of changes of mobility in Ge resulting from gamma irradiation with those resulting from neutron bombardment reveal the much more complex defect distributions existing after neutron irradiation (Cleland, et al., 1955a, 1956; Crawford, 1958; see also sec. 3.10). Cross sections for electron and hole capture, obtained from measurements of minority carrier lifetime in germanium and silicon, pro-

vide information on the charge held by a defect. The nature of complex recombination centers produced by bombardment has been unravelled through combined studies of lifetime, mobility, and Hall coefficient.

Finally, recent developments which highlight the important effects of impurity content on both the nature of defect states and the relaxation kinetics of radiation effects deserve mention. Brown, et al. (1959), have shown that the annealing of electron damage in n-type Ga depends markedly on whether the initial doping agent was phosphorus, arsenic, or antimony. The annealing which occurs in the range of 30°C to 100°C is markedly suppressed in specimens containing phosphorus and arsenic as compared to the behavior of antimony-doped specimens. Moreover, Wertheim and Buchanan (1959), in agreement with conclusions from electron spin resonance experiments of Bemski (1959) and Watkins, et al. (1959), have established that the level 0.16 eV below the conduction band in Si is associated with oxygen impurity.

To date, most investigations have been confined to the diamond structure semiconductors (diamond, silicon and germanium) and to the zinc blende intermetallic compounds listed in Table 3. There is a pressing need for extending these studies to less understood ionic and semi-ionic semiconductors such as Cu_2O , NiO , ZnO , TiO_2 and other oxides, sulfides, etc. The use of radiation for the introduction of defects may prove a valuable means of elucidating the role of defects in the electronic behavior of materials of this class.

Recently, work has been reported on the changes of lattice parameter and overall linear dimensions of semiconductors undergoing irradiation. Vook and Balluffi (1959a) measured changes of length in pure germanium single crystals irradiated at 25°K by 10.2 MeV deuterons and found $\Delta L/L$ to be $(1.5 \pm 0.3) \times 10^{-21}$ per deuteron per cm^2 . Less than 20% of this length change annealed out below 200°K; 50% below 300°K; and 85% below 360°K. Fujita and Gonser (1958) observed considerable small angle X-ray scattering in deuteron irradiated germanium single crystals at 130°K, much of which annealed out on warming to 300°K. Simmons (1959) measured a small increase in lattice parameter on the Vook and Balluffi specimen after it had been irradiated and then warmed to 320°K. Vook and Balluffi (1959b) discussed these results and attempted to reconcile the differences by proposing the existence of clusters containing interstitials and vacancies.

Gonser and Okkerse (1957), and later the Purdue group, observed large changes of volume in InSb and GaSb on deuteron irradiation at 150°K. The changes were much larger than those observed in germanium. It is possible that these differences arise from the fact that the binding in the III-V compounds is partially ionic, whereas that in germanium is wholly covalent. When an ionic crystal is disordered, large Coulomb energy increases produced by disordering might result in volume changes. It would be of value to calculate the Coulomb energy of a disordered and an ordered ionic lattice for various values of the lattice parameter.

Before closing this section, attention also should be directed to graphite, which is intermediate between the semiconductors and the metals. Because of its importance as a moderator in nuclear reactors, radiation effects in graphite have been the subject of intensive study and a large body of information has been accumulated. Most of the experiments have been performed with reactor irradiation. Electrical, mechanical, structural, and thermal effects, including stored energy and thermal conductivity, have been investigated, and also many annealing studies have been made. A good summary is given by Hennig (Harwood, et al., 1958, ch. VIII). Graphite is also of scientific interest because of its unusual atomic structure. Most of the observations have been explained in terms of lattice vacancies, interstitials, and clusters of interstitials. It is assumed that the interstitials lie between the atomic layers and are only weakly bound, giving rise to large expansion perpendicular to the layers and to large storage of energy. The complex and variable microstructure of graphite makes the experimental situation more complicated than with other well-documented materials.

2.5.3. Insulators

This enormous class of materials includes essentially all solids with the exception of metals and "good" semiconductors, running the gamut of chemical binding, including covalent crystals, and crystals in which the component atoms or molecules are held together by van der Waals forces; as well as gradations between these valence types. Excluded from consideration is the radiation chemistry of organic solids; discussion will be concentrated primarily on displacement effects in ionic, covalent, and partially ionic-partially covalent substances. Because of the wide variety of structures and compositions of crystalline solids falling within this group, it is difficult to organize simply and to coordinate properly the experimental investigations of radiation effects. It

growth of nuclei of precipitated vacancies. Nuclear magnetic resonance investigations also indicate the presence of molecular fluorine and metallic lithium in this exposure range (Ring, et al., 1958).

Two other observations, qualitatively different from results obtained with X-rays and gamma rays, have been made in KCl. Neutron bombardment causes as much as a 2-order-of-magnitude increase in the extrinsic ionic conductivity (Nelson, et al., 1953) which persists up to 170°C. In contrast, exposure to gamma rays and X-rays produces a decrease in this temperature range. Spaepen has observed a relative density decrease of about 0.1% after prolonged neutron bombardment ($\sim 10^{19}$ nvt) which appears to be more than two orders of magnitude larger than has been observed after X-ray exposure.

It seems increasingly evident that before displacement effects can be isolated for separate study, the extensive effects associated with photochemical processes induced by ionizing radiation must be more clearly understood. At the present time there are two proposed mechanisms for photochemical defect production: the Seitz mechanism, and the Varley mechanism (sec. 2.1). The numerous points of experimental evidence which favor each of these viewpoints will not be discussed here, but it must be noted that information is available to support each one. More work is necessary before the mode of defect production by ionizing radiation can be established conclusively.

Although color centers observed in neutron-irradiated alkali halides are qualitatively similar to those resulting from x-irradiation, marked differences have been observed in other materials. There are at least two neutron-specific absorption bands in Al_2O_3 (2040 Å and 2600 Å); the F-center (electron in an anion vacancy) which apparently is not produced by X-rays has been identified by electron spin resonance (Wertz, et al., 1957) in reactor-irradiated MgO; it has been shown that atomic displacements in diamond (Clark, et al., 1956) are necessary for the production of certain systems of optical absorption lines. Finally, it appears that certain bands (absorption complexes centered at $\sim 1650\text{Å}$ and $\sim 2200\text{Å}$), although produced in small yield by ionizing radiation, are associated with imperfections expected from the displacement processes in high-purity fused silica and crystalline quartz. The centers in quartz and fused silica have been investigated extensively by means of optical absorption (Mitchell and Paige, 1958), electron spin resonance (Weeks, 1956a, b), and magnetic susceptibility measurements (Stevens, et al., 1958). It appears that the

is no less difficult to summarize the widely scattered experimental studies. Therefore, in the following, an attempt will be made to focus attention on the more outstanding areas and experimental techniques where some degree of success has been attained.

Perhaps the largest body of experimental information relating to radiation effects concerns the area of alkali halides. However, most of these studies pertain to defect production by photochemical processes rather than to effects of direct atomic displacement. Indeed, a careful survey of studies employing energetic particles capable of producing atomic displacements reveals few qualitative differences in the physical behavior from those observed with ionizing radiation. Recently, careful studies of the kinetics of coloration of NaCl and KCl have demonstrated that lattice defects are produced by neutron bombardment and also, at a different rate, by gamma ray bombardment. These studies have enabled the production of defects to be distinguished from the coloration of existing defects (P. W. Levy, unpublished). The most extensive effects with fast particles have been observed (Binder and Sturm, 1955, 1957; Mayer, et al., 1956; Smallman and Willis, 1957; Gilman and Johnston, 1958; Lambert and Guinier, 1957a, b, 1958; Spaepen, 1958) in thermal neutron irradiation of LiF, in which the lattice damage is caused primarily by the energetic alpha particle and triton (total energy 4.8 Mev) released by the (n, α) reaction on Li^6 . Marked increases in lattice parameter ($\Delta a/a = 2 \times 10^{-3}$), increases in hardness, decreases in density ($\Delta \rho/\rho$ up to -22%), and intense coloration have been observed. X-ray studies on specimens both after irradiation and after post-irradiation annealing (above 300°C) indicate three ranges of behavior: (a) a linear increase in lattice parameter, up to $\sim 10^{17}$ nvt exposure, which can be recovered upon subsequent annealing, (b) a progressive decrease in $\Delta a/a$ with further exposure (up to $\sim 2 \times 10^{18}$ nvt) accompanied by evidence of shear strain in X-ray patterns and a large increase in small-angle scattering (annealing removes lattice strain but not all of the density decrease of small-angle scattering), and (c) at higher exposures a progressive removal of evidence of strain and restoration of lattice parameter and not much additional change in either small-angle scattering or density. This behavior is interpreted as indicating (1) the steady introduction of interstitials and vacancies which expand the lattice up to $\sim 10^{17}$ nvt, and (2) the coalescence of defects into platelets at higher doses. After large exposures ($> 10^{17}$ nvt), annealing in the 400°C temperature range causes the appearance of microscopically visible cubic voids which apparently result from the

1650Å absorption complex is associated with holes attached to interstitial oxygen ions and that the 2200Å complex is caused by electrons in the sites of missing oxygens.

The most extensive effort in this area has been devoted to studies of structural effects of fast neutron bombardment, which are summarized in Table 7. Investigations of other defect-sensitive properties also have been made on neutron-bombarded specimens. Low temperature thermal conductivity measurements (Berman, 1951; Berman, et al., 1955) on quartz, Al_2O_3 , and diamond reveal a marked decrease in this property which presumably results from the scattering of lattice waves from point defects and disordered regions introduced by the bombardment. Specific heat measurements on neutron-irradiated quartz indicate a temperature dependence below 50°K of the type observed in fused silica (Westrum, 1956). The marked decrease in hardness of irradiated diamond (Damask, 1958) is correlated with the decrease in cohesive energy expected from the decrease in density.

2.5.4. Materials Suffering Transmutations

Since each element which suffers a nuclear reaction occurs in a large variety of solids, it is difficult to generalize the effects, particularly since each element reacts differently with different bombarding particles of different energies. Although some attempt has been made to study each type of effect, past effort has been primarily concentrated upon materials and environments that are of practical importance. Consequently, the behavior of uranium and uranium alloys and compounds during slow neutron bombardment has received most emphasis. Again, for practical reasons, attention has been concentrated upon the behavior of the inert gas fission products, Xe and Kr, because these gases produce, at high temperatures, either large volume increases or disintegration. Thus, the behavior of these transmutation products within the solid at various temperatures, and their release as a function of temperature (Barnes, et al., 1958) have been studied.

Much of this work has been of an exploratory nature, using impure materials in a wide variety of initial conditions and in complex environments. Uranium, and many of its alloys (Zr, Mo, Nb) and compounds (oxides and carbides) have been explored most thoroughly at various temperatures in reactor neutron fluxes. In a few experiments the heating has been performed after low temperature irradiation. Boron (Pease, 1954), beryllium (Ells and Perryman, 1959; Rich, et al., 1959), and lithium (Lillie, 1958) and their alloys and compounds also have been irradiated in re-

actors. Usually density change was measured in conjunction with microscopic examination.

There are many practical difficulties in studying such classes of solids. Normally, high doses are necessary (isotopic enrichment is often employed) and the samples are frequently very radioactive. The energy release during the transmutation can be large, making the control of temperature during irradiation difficult; it is particularly difficult to perform low-temperature experiments on large samples.

Because of the high kinetic energy of the particles after transmutation, very intense local displacement damage occurs. Although displacement effects in uranium have not been studied as intensively as in other materials because of the practical difficulties and greater complexity of interpretation, certain aspects, significant again because of their practical importance, have been studied, i.e., the radiation-induced growth (Pugh, 1955; Foote, 1955; Paine and Kittel, 1955) and creep (Roberts and Cottrell, 1956) of uranium, and the homogenization produced by fission

TABLE 7
STRUCTURAL EFFECTS IN NON-METALS CAUSED BY NEUTRON BOMBARDMENT

Material	Nature of effect	Comments
Diamond	Lattice expansion	$\Delta \rho / \rho$ saturates at -4% for low dose rate; at high dose rate structure becomes amorphous.
Quartz	Large density decrease; vitrification	Saturation $\Delta \rho / \rho = -15\%$ at dose 1.5×10^{20} nvt.
Fused Silica	Density increase	Saturates at same value of 7 as attained by quartz.
Coesite (High density form of SiO_2).	No detectable change of lattice parameter.	
Zircon	Lattice expansion (anisotropic) Structure retained.	No evidence of metamict state that is found in nature.
Beryl	Density decrease; vitrification	$\Delta \rho / \rho = -6\%$ at saturation (3.6×10^{20}).
Barium Titanate	Expansion; transformed from tetragonal to cubic.	Cubic phase has unusual thermal stability. Tetragonal form cannot be recovered by annealing.
Zirconia	Expansion, transformed from monoclinic to cubic.	Transformation only observed in specimens with uranium impurity; monoclinic form recovered on annealing above 800°C.

fragments in uranium-molybdenum and uranium-niobium alloys (Bleiberg, et al., 1956; Konobeevsky, 1956).

3. NEEDS AND PROSPECTS

In this section we propose to discuss a rather extensive set of problems that require solution and in which substantial progress seems attainable in the near future. We shall also discuss areas in which new experimental techniques are sorely needed and instances where techniques already exist but are not being applied as extensively as they deserve. Many of the prospects for immediate advancement in the field will become clear from this section. However, this should not be considered as a blueprint for the future development of the field. The strikingly novel ideas that sometimes advance a subject most dramatically are unlikely to be forecast in a list of this type.

3.1. Development and Use of New Observational Techniques

The damage produced by bombarding solids with atomic particles is itself atomic in scale. If radiation effects are to be understood sufficiently so that confident predictions may be made, response of materials to radiation environments must be understood on the same scale. Most of our present knowledge of the atomic behavior is either abstractly concocted from what little we know of atomic forces, or else deduced from the effect that radiation exerts on some bulk physical property, which is very often itself little understood in atomic terms.¹

The early information on dislocations was derived in a similar way, but now these methods have been superseded by those involving direct observation. Similarly, direct observation of the defects produced by radiation will increase our understanding. Not only can the crystallography of the defect clusters be determined, but it may be possible to determine the number and behavior of point defects from which they are composed.

3.1.1. Etching

Observations on lithium fluoride (Gilman and Johnston, 1956; Whapham, 1958) have shown some of the advantages of directly observing the behavior of dislocation lines, as revealed by an etch-

¹ The difficulty in interpreting the atomic behavior underlying changes in the electrical resistivity caused by annealing irradiated material (usually labeled Stage I, II, etc.) illustrates this.

ing technique. Such techniques (either chemical or physical) for removing surface atoms can reveal the presence of voids (Barnes, 1952), gas bubbles (Barnes, et al., 1958), inclusions, dislocation lines, spikes (Young, 1958), etc. If, as is sometimes necessary, impurities must be associated with imperfections before they can be suitably etched, then the scope of the method is restricted to observations on those defects which exist after the necessary aging treatment.

3.1.2. Thin Films

Thin film electron transmission microscopy offers the high resolution and magnification which are necessary to observe defects down to about 20\AA in size. Inasmuch as the defect itself is the cause of the diffraction of the electrons, there is no necessity for any artificial enhancement for the observation of defects, although any segregation associated with a defect normally can be inferred from the diffraction characteristics. This technique already has been used to show the nature of vacancy agglomerates and their annealing characteristics in quenched metals (Hirsch, et al., 1958). The behavior of the vacancies themselves can be inferred and their numbers estimated. The effects in irradiated material, although more complex, should be similar to those in quenched materials, and the first observations of this character have been reported (Wilsdorf, 1959). Changes of dislocation lines caused by radiation can be observed in this way, and disordered regions, spikes, fission tracks, etc., should be detectable if the resolution is adequate. Electron diffraction patterns from regions containing the defect permit the crystallography of the defects to be determined. The development of dark field techniques would assist in discernment of defects without excessive heating in the electron beam.

There are several limitations to this method, however, which should be mentioned.

(a) As with most microscopic methods it is difficult to compare slight changes, but such changes can be studied quantitatively in other ways once their atomic nature has been established.

(b) If thin films themselves are irradiated, the close proximity of the surface (which is a defect sink) may radically alter the effects from those which occur in bulk material. However, the material may be thinned after bombardment.

(c) Surface effects on chemically active materials can be troublesome, but techniques to protect the surface can be developed, e.g., thin protective layers of plastic may be deposited.

(d) Low temperature studies are difficult, but low temperature techniques, i.e., physical etching and cold stages, should be developed. Until this is achieved, high-melting-point materials can be investigated to obtain an indication of low-temperature defect-configurations.

3.1.3. Moiré Patterns

The transmission of electromagnetic waves, of wavelength less than the lattice repeat size, through two superimposed films produces moiré patterns, if the materials have different repeat distance or are rotated relative to one another (Menter, 1958). The superposition of a thin film of defect-free material upon one which has been irradiated should reveal, in an electron microscope, defects, whether they be dislocations or disturbed regions. Thus, the moiré patterns can be used to measure changes in repeat distance, misorientations, etc., and should be particularly useful for studying spike defects in detail.

3.1.4. Field Emission Microscopy

The field emission microscope has been used to study sputtering caused by bombarding particles, which is closely allied to radiation effects (see also sec. 4.). Field emission microscopy can be a powerful tool for radiation studies because of the opportunities to observe single atoms. Although only surface atoms can be "seen," disarrangements or defects within a solid should be detectable after removing, *in situ*, layers of atoms by sputtering.¹ The device may be particularly useful for observing low temperature effects if samples can be cooled sufficiently.

3.2. Experiments with Monoenergetic Neutrons and with Very High Energy Neutrons and Other Particles

It is unfortunate that the energy spectrum of reactor neutrons cannot be conveniently predetermined or measured, as is the case for electron, alpha particle, or deuteron beams from ion accelerators. The best definition attained in a reactor is by means of fission converters. Even with these the spectrum is quite wide (0.1 Mev to ~15 Mev) and it is difficult to exclude epithermal neutrons originating outside the converter. Therefore, most of the available

¹ This type of physical etching might be developed for the examination of surfaces and foils in the electron microscope. It has the added feature that it can be carried out at low temperature.

information on fast-neutron-induced lattice damage has been obtained with a broad spectrum of neutron energies. Nevertheless, it is evident that experiments employing monoenergetic neutrons are urgently needed for establishing several important points concerning the basis of fast-neutron damage. It is desirable to have reliable measurements of the number of lattice displacements per collision as a function of the energy of the incident neutron for comparison with the values predicted by theory.

A start in this direction has been made by Ruby, et al. (1958), who employed a Van de Graaff accelerator to produce neutrons in the range 1.7 to 4.8 Mev from the $H^2(d,n)He^3$ reaction for studies of effects in high purity germanium. Also, Curtis and Cleland (to be published) have used 14 Mev neutrons from the $H^2(d,n)He^4$ reaction for studies of the same material. Unfortunately, the available neutron fluxes from existing machines (Van de Graaff, Cyclotron, Cockroft-Walton, etc.) are quite small, and materials with properties sufficiently sensitive to the small concentrations of damage obtainable are extremely limited. To increase neutron fluxes it is necessary to find targets capable of retaining higher concentrations of deuterium or tritium at the high temperatures produced by intense beams of ions. Developmental work in this direction would be desirable.

Damage obtainable with 14 Mev neutrons (as from the (d,t) reaction) may differ in some cases from damage occurring in reactors. Preliminary work has not yet shown much difference; however, the experiments thus far have not been extensive. In materials of low and intermediate atomic weight, raising the bombarding energy from 1 Mev to 14 Mev simply causes the most energetic knock-on atoms to be born in the higher ionization range (sec. 2.1). Thus more ionization will be produced, but not much more displacement damage. For substances of high atomic weight, however, 1 Mev neutrons create knock-ons which are well below the ionization range. Increasing the energy of the neutrons to 14 Mev causes the most energetic knock-ons to occur much higher up in the displacement range of energies, and should increase the severity of any displacement spikes that occur. Since displacement spikes are also expected to be more important in elements of high mass, spike phenomena may be observable to a much greater extent. An experimental difficulty at present is the low flux of high-energy neutrons that can be obtained. The practical importance of such effects, if they exist, in the technology of controlled fusion is obvious.

A small amount of work has been done with extremely high-

energy heavy-charged particles. Smoluchowski (1956) and Kobayashi (1956, 1957) have irradiated alkali halides with 400 Mev protons from a cyclotron. Nuclear reactions play an important role in these experiments, and the products of the reactions bombard the specimen internally. It is interesting to note that the range of a proton of such high energy is several centimeters, and thus real volume effects are obtainable and specimens of convenient thickness can be used; fine wires and thin films are not necessary. This advantage of high penetration also is shared by gamma rays and by neutrons, but the damaging power of the former is weak, and neutrons have the disadvantage of either possessing a broad spectrum or being of low intensity. Greater use of very high energy protons or deuterons seems indicated.

3.3. Invention of New Methods of Distinguishing Between Defects, Particularly Interstitials and Vacancies

Experiments should be done on simple metals that will distinguish between the presence of interstitials and the presence of vacancies. Shockley (see Fisher, et al., 1956, p. 581) suggests that such differentiation may be made by noting the direction of climb of a small-angle grain boundary when the defect in question anneals out of a specimen. For a second method see section 3.12 below. Proof that the defects in aluminum at high temperatures are vacancies is presented in section 3.26. We should be able to decide on the basis of recent transmission electron microscope experiments whether the small plates formed are interstitial platelets or collapsed vacancy platelets, since the resulting stacking faults differ. Moreover, interstitials and vacancies will cause a given screw dislocation to spiral in opposite directions. Still another method might be founded on the fact that substitutional impurity atoms smaller than atoms of the matrix should tend to trap interstitials in preference to vacancies, while impurity atoms that are larger should tend to trap vacancies. In non-metals nuclear magnetic resonance (Hon and Bray, 1959) can be used.

3.4. Wide Use of Radiation-Enhanced Diffusion

The rate of self-diffusion is proportional to the number of mobile point defects present, and any treatment, e.g., quenching, deformation, or irradiation, which increases this number will enhance the diffusion rate. The enhanced diffusion produced by radiation is of practical importance in reactor materials, e.g., effect on creep rates and precipitate coarsening rates during irradiation,

and can provide information on the numbers of point defects produced, their mobility, and the nature of their eventual sinks.

Any property which depends on diffusion may be used to "observe" such effects, but unless the rate of production of defects is high and they are able to make a large number of jumps before capture or annihilation, the diffusion will be on a fine scale and must be looked for on this scale. Thus Lomer (1954) estimated that only diffusion over distances less than $\sim 10^{-6}$ cm would be observed after neutron irradiation of copper. The effect may be transient as after quenching, or prolonged, e.g., after reaching equilibrium during creep or irradiation (Dienes and Damask, 1958; Wechsler and Kernohan, 1958; Thomas, 1956). Although large numbers of vacancies can be quenched into materials (to a maximum atomic concentration of about 10^{-4}), the diffusion process will be complicated by agglomeration of the vacancies (which occurs with such high concentrations) and by the close proximity of the sinks so produced, which limit the jumps by the vacancies and thus any anticipated enhancement of diffusion. The rate of production of point defects by steady deformation may reach 10^{-7} sec.⁻¹, and by neutron radiation from a reactor may be as high as 10^{-9} sec.⁻¹. Much higher values, however, can be obtained from other artificially accelerated particles. Comparisons of the enhancement produced by deformation and radiation should be useful.

The temperature dependence of diffusion enhancement may indicate the nature of the defect annihilation process and in certain instances the defect involved. At lower temperatures there may be discontinuities in the enhancement due to a restriction in the movement of the defects, recombination becoming more important. Eventually, the temperature at which the enhancement ceases will indicate the mobility of the defect responsible.

The relative increases in diffusion rates of materials differing in structure or purity but exposed to the same radiation environment, should provide information on the nature of sinks. Comparison of the effects produced by various radiation environments on a given material can reveal relative rates of production of point defects. Also, the absolute value of the radiation-induced defect concentration can be determined if the thermal point-defect concentration is known.

A review of all the information on the effect of radiation on diffusion-type processes would be useful.

3.5. Studies of Transient Damage with Accelerators and Pulsed Reactors

A few pulsed reactors that give intense bursts of fast neutrons for brief periods have been constructed. An exposure of the order of 10^{14} fast neutrons per cm^2 can be obtained in a single burst, with a duration of a hundred microseconds. Such an exposure is sufficient to affect electrical properties of semiconductors, optical properties of insulators, and the ultrasonic response of many substances. It thus becomes possible to study small amounts of damage produced at rates 4 to 6 orders of magnitude faster than in ordinary reactors, and it is quite possible that unexpected phenomena will be found. It is of interest to know the extent to which a reciprocity law applies to radiation damage, that is, whether a given total dose applied quickly produces the same effects as the same dose applied slowly. If no recovery or rearrangement occurs during irradiation, it might be expected that reciprocity would obtain. This then provides a method of testing whether recovery or annealing is taking place. Customarily, in order to determine whether annealing has occurred during an irradiation, exposure is repeated at a lower temperature and the two results are compared. For metals, experiments conducted at continually decreasing irradiation temperatures have revealed a steady succession of new annealing stages, until the lowest temperature which can be maintained in a reactor, namely about 4°K , has been reached. It is important to know if annealing stages occur below this temperature, and a test of the reciprocity law at low temperatures might be most suitable for obtaining such information.

The recovery kinetics following pulsed irradiations could be observed on a millisecond time scale, which would be 3 to 6 orders of magnitude smaller than the times involved in conventional recovery experiments. It would be interesting to study transients in this range, to see if unexpected effects can be found. An important new variable which would assist in unravelling complex annealing kinetics might be provided by such experiments. It is also possible to perform pulsed irradiations with certain types of accelerators—a tool which thus far seems to have received little attention.

3.6. Wider Application of Magnetic Resonance Techniques

Paramagnetic resonance has been a revolutionary tool for the study of imperfections in solids. When properly applied to appropriate systems, the techniques of both electron spin resonance

(ESR) and nuclear magnetic resonance (NMR) are capable of yielding detailed information pertaining to the environment of a specific defect. Recently, ESR has been applied with varying degrees of success to a number of irradiated crystals, among which are diamond, quartz (Griffiths, et al., 1955), fused silica (Weeks, 1956), magnesium oxide (Wertz, et al., 1957), and silicon (Schulz-Dubois, et al., 1955). The recent discovery by Bemski (1959) and Watkins, et al. (1959) that one type of irradiation-induced defect in electron-bombarded silicon, previously thought to be basic to the damaged lattice, is in fact a complex of a defect and an oxygen impurity atom, is an excellent example of the versatility of this technique. In another area, Ring, et al. (1958) have shown, by means of NMR, that thermal-neutron-irradiated ($>10^{17}$ n/cm²) LiF contains both molecular fluorine and metallic lithium. It is expected that future application of resonance techniques to studies of radiation effects in solids will be exceedingly fruitful in establishing the detailed nature of lattice defects. It is very likely that electron-nuclear double resonance (ENDOR) methods, which have been so fruitful in investigations of F-centers in alkali halides (Feher, 1957), also will find application.

3.7. Wider Use of the Slow-Neutron-Transmission Technique

The absolute number of displaced atoms in a material can be determined, if the concentration is greater than about 10^{-5} , by studying scattering of slow neutrons. This may be done by measuring either the attenuation of the transmitted neutrons or the intensity of the scattered ones. The scattering due to vacancies and interstitials cannot be discriminated, and relaxations around the defects can modify their contribution. However, the interpretation of the results is less ambiguous than that of the scattering of conduction electrons or the scattering of X-rays. The state of agglomeration of point defects can be inferred by studying either transmission as a function of neutron wavelength or angular distribution of scattering at a constant wavelength. Only agglomerates smaller than $\sim 40\text{\AA}$ can be detected, and this limit can lead to uncertainty in estimating the number of defects present. However, agglomerates above this size should be discernible in the electron microscope, which should be used in conjunction wherever possible, particularly as only then can the crystallography of the large clusters be determined.

A study of the scattering of X-rays also could be used in conjunction, although neutron scattering has several advantages. If the neutron wavelength is longer than the Bragg cut-off, then the

neutrons are scattered only by lattice defects (neglecting inelastic, isotope, and spin effects, which can be determined separately with unirradiated but otherwise similar material). X-rays, on the other hand, are highly absorbed, and double Bragg reflections can mask the effects so that thin single crystals must be used.

The slow-neutron technique has been used to study the atomic defects in graphite (Antal, et al., 1955), quartz (Mitchell and Wedepohl, 1958), and sapphire (Antal and Goland, 1958), after neutron irradiation, and in copper and aluminum (Atkinson, 1958) after deformation. Only materials containing elements of low absorption and low incoherent scattering cross sections can be used, e.g., Be, C, N, O, F, Mg, Al, Si, P, S, Pb, Bi. The limitation to the sensitivity of the method is the available intensity of neutrons of wavelength greater than the Bragg cut off. Thus, materials with small repeat distances present an advantage because the intensity of slow neutrons from a reactor varies as λ^{-5} . Neutrons can be obtained from a high-flux reactor and moderated in the reactor with a material at very low temperature. This would probably increase the intensity by more than an order of magnitude over previous levels.

Because of the lack of intensity, large samples (greater than one-inch cube) are normally necessary, and unless very high concentrations of defects are present, the radiation damage must be distributed throughout the sample. Only the radiation effects due to neutrons and γ -rays can be usefully studied.

The technique is particularly suited to the determination of the numbers of point defects introduced by irradiation and to a study of their subsequent clustering and annihilation by annealing. Ideally the irradiations should be performed at very low temperatures and the samples should be annealed in stages until no defects are observed. Such experiments are more difficult than those involving irradiation and examination at room temperature. The use of high-melting-point materials should keep the experiment flexible and yet give a full indication of displacement and annealing processes. Irradiation at low temperature and examination at room temperature may help to obtain positive results where the number of defects produced by room temperature irradiation is near the limit of the technique. A study of the numbers of vacancies introduced by quenching and of their subsequent clustering would assist the interpretation and understanding of the more complex radiation processes.

3.8. More Use of Electrons and Gamma-Rays and Intercomparison of Results Obtained with Various Kinds of Radiation

Until 3 or 4 years ago, the use of electron irradiation for damage studies was limited principally to one application, the determination of threshold displacement energies. More recently it has been used as a means of producing Frenkel defects in copper, nickel, germanium, silicon, and InSb, in order to study their recovery behavior and their effects on electrical properties, mechanical properties, and stored energy. The number of such studies has been limited, but, as a general rule, the results obtained have been unusually informative. Modulus measurements indicate, for example, that point defects are capable of pinning dislocations (Dieckamp and Sosin, 1956), whereas the dislocation pinning resulting from neutron irradiation (Thompson and Holmes, 1956), while interpreted as the result of point defects, could not unambiguously be proved to be such, since neutron irradiation may produce other kinds of lattice damage in addition to point defects. Furthermore, the recovery of damage introduced by electron irradiation is generally found to closely approximate "ideal" types of kinetics, whereas heavy particle irradiation usually leads to more complex recovery kinetics. In view of such advantages and the fact that the models for production of damage by electron irradiation are better defined than for heavy particle irradiation, it is important that more extensive use be made of electron irradiation as a means of introducing lattice defects in solids.

Similar advantages over heavy particle irradiation are presented by γ -radiation. In addition, there is the advantage that larger specimens may be irradiated, because of the greater range in solids of γ -rays as compared with the range of electrons. The cross sections for producing atomic displacements by γ -rays are in general smaller by a factor of the order of 100, than for electrons of equivalent energy, the principal contribution coming from Compton recoil electrons, which in turn displace atoms (Dienes and Vineyard, 1957; Owen and Holmes, 1959). The use of γ -radiation is, therefore, of most value in experiments in which only a small concentration of point defects is required to produce a significant property change. Phenomena which have been produced and studied with γ -radiation include ordering of Cu_3Au (Dugdale, 1956), carrier removal in semiconductors (Cleland, et al., 1956), and internal friction changes in pure metals (Thompson and Holmes, 1957).

3.9. Careful Annealing Studies on Very Pure Irradiated Specimens

More annealing studies are needed with copper and silver, and, in particular, the annealing of irradiated gold should be studied in the range from 4°K to 80°K. All three metals should be studied in the irradiated and in the quenched state from 200°K to 520°K. At present, good annealing data exist for quenched but not for irradiated gold. Data on high-purity copper and silver does not exist in the range from 220°K to 520°K, after either irradiation or quenching. It also would be of value to investigate the annealing behavior of pure metals which had been irradiated after the introduction of quenched-in vacancies.

Our knowledge of the interaction of defects at present is most rudimentary. It has been suggested that point interstitials can attract one another, but no experimental proof is available. In this connection it might be possible to perform relaxation experiments on irradiated specimens after various anneals. A di-interstitial would have an axis and probably would be oriented by stress, thus giving rise to internal friction phenomena.

Palmer and Koehler (1958) have performed some experiments with gold which indicate that the binding energy of the divacancy is low (i.e., $B < 0.15$ ev). DeSorbo and Turnbull (1959) have reported on experiments which indicate an appreciable binding energy of a divacancy in aluminum.

Thus far, not enough attention has been given to the possible interaction of the defects with impurity atoms. Blewitt and co-workers (1957) have shown that this can be important in low temperature irradiation studies. DeSorbo and Turnbull have shown that it is important in quenched aluminum.

3.10. Studies of Directional Effects

In almost all reactor irradiations the neutron flux is essentially isotropic, and the damage which results shows no directionality except that associated with the crystalline axes of the specimen. By placing a fission converter near the specimen, it is possible to introduce a strongly directional component of the fast neutron flux. Truell, Teutonico, and Levy (1957) have used this arrangement to bombard a silicon single crystal along one of its cubic axes, and have found that the crystal is made acoustically birefringent by this treatment. This provides evidence that the defects produced by a fast neutron are not distributed in a spherically symmetrical pattern, but in an asymmetric way related to the direction of incidence of the neutron. From theoretical models

of the damaging event, the pattern would be expected to be rather cigar-shaped, with its long axis parallel to the direction of incidence. The experimental results are consistent with this interpretation. Since only one physical property has been checked, and this in only one substance, it would appear that a large field of investigation remains open. Electrical and optical properties may also show directional effects and many other substances should be investigated. More detailed knowledge of the shape and size of the elementary damaged regions should be the fruit of such work.

In charged particle bombardment, the incident beam is almost always well collimated just as it comes from the accelerator. However, the penetration depth of charged particles of the usual energy is so small that thin foils or wires are used as targets and any directionality of the damage in these specimens is difficult to observe. Owing to the nature of the Rutherford law, the knock-ons resulting from heavy, charged-particle bombardment are found in the forward direction less often than those resulting from neutron bombardment. Nevertheless, directional properties of the damage produced in such irradiations would be expected, and it would be worthwhile to devise experimental means for their observation. For example, a whisker of a transparent cubic crystal with its long axis in the x -direction, bombarded with charged particles in the y -direction, may become birefringent to light passing in the z -direction.

When collimated radiation falls upon a single crystal, the effects induced may depend on the orientation of the radiation with respect to the crystalline axes, even though a scalar property is being measured. This kind of directional effect should be particularly prominent in the case of electron bombardment at energies near the threshold energy for producing displacements, since all considerations indicate that the probability of displacing an atom depends very much on the direction in which it is driven. Pronounced directional effects of this type have been observed in the irradiation of germanium (Brown and Augustyniak, 1959). Probably the only reason that such effects have not been reported in other substances is that they have not been looked for. In connection with the more detailed inquiries now being made into the displacement of atoms at low energies, directional observations near threshold would be very helpful. It would be particularly useful at this stage to have such studies made for copper (at low temperatures) and for other close-packed metals. The suggestions for electron microscopy of thin foils discussed in section 3.1 are also pertinent.

3.11. Studies of Chemical Effects in Solids

Displaced electrons and displaced atoms may lead to a variety of chemical reactions, particularly in covalent solids, via bond breakage and reformation. Decomposition, gas formation, degradation, cross-linking, and polymerization are typical examples of such reactions. These topics belong properly to radiation chemistry rather than solid-state research and will not be treated any further. There is an important class of phenomena, however, which is at least partly the province of solid-state research, namely, the influence of nuclear radiation on the chemical activity of solids. There are a few examples in the literature indicating that these effects can be very important indeed. Some representative studies are outlined below.

1. Systematic investigations have shown (Hennig, et al., 1958; Kosiba and Dienes, 1957) that prior irradiation of graphite in a reactor increases its subsequent thermal oxidation rate by about a factor of six at 300°C (decreasing to about a factor of two at 450°C). This effect does not depend on the impurities present in reactor grade graphite (very low impurity level). The increased reaction rate persists even after 25% of the sample has been oxidized. Thus, the displaced atoms exert a catalytic influence on this heterogeneous solid-gas reaction.

2. Irradiation with high energy protons alters the electrochemical properties of tungsten as well as the rate of solution of Fe_2O_3 in HCl (Simnad and Smoluchowski, 1955a, b). Chemical reactivity is increased in both instances. A systematic study of these and similar reactions has not been reported.

3. The reaction between NiO and H_2 is greatly influenced by prior high energy proton irradiations (Simnad, et al., 1958). The induction period is reduced and the rate of reaction is increased. Other rather cursory reports in the literature indicate that in some cases compound formation occurs on the surfaces of crystals upon irradiation (Bloch, 1953; Wieninger and Adler, 1956).

4. The alteration of catalytic activity by gamma irradiation was investigated in a more direct manner in two recent studies. Taylor and Kohn (1957) found that the hydrogen-deuterium exchange reaction at -78°C was much faster over gamma alumina irradiated with gamma-rays (at -78°C) than over the unirradiated catalyst. Maurin, Balantine, and Sucher (1959), studied the ortho-para hydrogen conversion over several metal oxides irradiated with gamma rays. The rate of this reaction was substantially (20 to 80 fold) increased when the reaction was carried out over irradiated gamma-alumina, alumina-molybdena and reduced

molybdenum oxide, provided the irradiations were carried out under an atmosphere of purified hydrogen. Gamma ray irradiations in vacuum induced no catalytic activity.

5. The response of irradiated materials to chemical etching has been found to differ from that of unirradiated materials. Effects have been observed in germanium and silicon (Chang, 1957; Noggle and Stiegler, 1959), in copper (Young, 1959), and in lithium fluoride (Young, 1958). Etching effects are also considered in section 3.1 above.

It is clear from these examples that interesting avenues of research have been opened up. It is also clear that our knowledge is extremely fragmentary and interpretation is as yet non-existent. Radiation can be considered as a very important tool in this field, since large concentrations of electronic and atomic defects can be introduced in a rather well-controlled manner. Some rather obvious, although by no means easy, basic investigations may be suggested, as follows:

1. Compare the adsorption and desorption behavior of inactive (physical adsorption) and active (chemisorption) gases over unirradiated and irradiated crystals. The effects of ionizing radiation and fast-particle irradiation should be compared.

2. Conduct careful and systematic studies of simple gas reactions over unirradiated and irradiated crystals, e.g.:

- (a) Compare the influence of gamma irradiation with reactor irradiation, to determine role of displaced electrons and displaced atoms.

- (b) Investigate the role of gaseous environment during irradiation.

- (c) Characterize the radiation effects in solids by some physical property change and correlate with chemical effects.

- (d) Compare chemical results with adsorption and desorption measurements.

3. Follow up the observation that the electrochemical potential of a metal can be altered by fast-particle irradiation.

- (a) Is this a bulk effect, i.e., can it be characterized by the increase of the free energy of the material?

- (b) Is it a surface effect? Study equilibrium and non-equilibrium electrolytic processes.

4. Compare in a systematic way the dissolution of irradiated and non-irradiated solids. Gross chemical and etching experiments are indicated.

3.12. Further Studies of Effects on Elastic Constants

Dienes (1952) and Nabarro (1952) discussed the contribution of interstitials and vacancies to the elastic constants of a close-packed metal, in particular copper. They both concluded that an interstitial would increase the elastic constants whereas a vacancy would decrease them. Zener (1949) claimed that the elastic constants would decrease in either case. Thompson, Blewitt, and Holmes (sec. 2.5) found no change in a copper single crystal during a week in the pile at 18°K. This experiment was performed after the dislocations were completely pinned. Dieckamp and Sosin (1956) found a decrease in Young's modulus of copper in a similar experiment using electron irradiation at 78°K. The effect in an alloy, where the dislocations are pinned at the outset, should also be investigated.

3.13. Further Study of Effects of Radiation on Plastic Properties

Studies of the effects of radiation upon the mechanical properties of materials involve mainly the influence of radiation upon the behavior of the glissile dislocation lines. Their behavior will be influenced by large spikes (which might even create new lines) and by interaction with the point defects produced by the radiation. These defects can (a) pin the dislocation lines (causing dislocation hardening) and (b) hinder their flow (causing lattice hardening). There are several ways of distinguishing between these two effects. At low strains, as in internal friction studies, only the former type of hardening is observed. A study of the lower yield stress (σ_y) as a function of the grain size (d) can separate the two components K_y (due to dislocation hardening) and σ_1 (due to lattice hardening), if the analysis of Petch (1953) is assumed.

In this analysis

$$\sigma_y = \sigma_1 + K_y d^{-1/2} \quad (3.1)$$

By studying the dislocation and the lattice hardening separately, the defects on the dislocation lines can be distinguished from those in the bulk lattice (Adams and Higgins, 1959). Thus, the movement of the defects from the body of the crystal to dislocation lines during the annealing processes can be investigated after low temperature bombardment. The simultaneous study of other physical properties, e.g., electrical resistivity, should assist interpretation. Bombardment with electrons,¹ which even at liquid nitrogen

¹ γ -radiation produces homogeneous electron damage in metals, but the effects with attainable source strengths are much less than with electrons (sec. 2.1.).

temperature still hardens copper, would avoid complications arising from spike formation. This type of an investigation involves the use of thin wires or foils; however, the development of techniques for measuring the mechanical properties of thin samples would enable other nonpenetrating radiations to be studied and the effects of spikes to be elucidated. Such techniques should be developed.

This ability to distinguish defects on the dislocation lines from those in the crystal should be exploited. For instance, the competition of sinks other than the glissile dislocations can be studied; the distribution of defects, after irradiation at low temperature and subsequent annealing, can be compared with that after an equivalent irradiation at the temperature of the anneal; also the distributions after various types of bombardment can be compared.

The effect of radiation on the mechanical properties of non-metallic materials is complicated by the effect of ionization, but these materials have several practical advantages over metals, e.g., the comparative ease with which dislocation lines can be seen. Work reported by Johnston and Gilman (1959), in which the velocity of an isolated dislocation was measured as a function of applied stress in pristine and irradiated states, is especially valuable and should be pursued vigorously.

The body-centered cubic metals should be studied to elucidate the effect of crystal structure on such features as point defect mobility (crowdion behavior) and the limitation of ductility.

3.14. Critical Studies of the Mechanisms of Ordering and Disorder by Radiation

For alloys exhibiting order-disorder transformations, heavy-particle irradiation can result in either an increase (Blewitt and Coltman, 1954; Brinkman, et al., 1954) or a decrease (Siegel, 1949; Aronin, 1954) in degree of order, at temperatures where the thermal transformation rate is negligible. Two separate effects are involved. If the temperature of the specimen is maintained sufficiently high so that the vacancies produced by the irradiation are mobile, their migration results in a rearrangement of atoms into a configuration in which the free energy of the crystal is decreased. This can result in an increase in the degree of order. The disordering effect, on the other hand, is at least approximately temperature-independent, and is the result of the production of a different kind of lattice damage. The two effects can be separated by performing equivalent irradiations on similar

specimens held at different temperatures. If resistivity is used to measure the degree of order, the resistivity change which results from the production of Frenkel defects must be distinguished from that which results from changes in the degree of order. Since the resistivity contributions from defects and from disorder are additive, this separation can be achieved by comparing the rate of change of resistivity of ordered and disordered specimens subjected to similar irradiation.

The disordering of ordered alloys by heavy-particle irradiations has not been investigated in sufficient detail to enable an unambiguous assignment of a particular mechanism to this phenomenon. It seems likely that more than one mechanism may be operative and that the kind of irradiation, and the ratio of the masses of the atoms which comprise the alloy may determine the relative importance of the possible mechanisms. It appears likely that displacement spikes are the dominant mechanism responsible for disordering of Cu_3Au by protons and α -particles (sec. 2.2.4.). Since the ratio between the disordering rate of Cu_3Au by protons to that by neutrons is correctly predicted by the displacement-spike model (Brinkman, 1956a), it is likely that this mechanism is dominant also for neutron irradiation. Replacement collisions can account for the disordering of Ni_3Mn by neutron irradiation, but the disordering rate of this alloy by charged-particle irradiation has not been measured. Displacement-spike models and replacement-collision models will in general predict different ratios for the disordering rates produced by neutron and proton irradiation. To determine the dominant mechanism in Ni_3Mn , therefore, the disordering effects of charged particles need to be measured. Generally, it seems desirable to measure the disordering rates of more ordered alloys as a function of radiation environment. High energy electron irradiation (approx. 3 Mev) should also be included, since it should be capable of producing appreciable numbers of replacement collisions but no displacement spikes. The degree of order should be followed by monitoring more than a single physical property wherever possible, in order to determine changes in both short and long range order.

3.15. More Measurements of Stored Energy

One of the most significant measurements that can be made, particularly with the noble metals, is the release of stored energy. It generally has been possible to make relatively more accurate calculations of the self-energy of lattice defects than of their contribution to, say, electrical resistivity. Furthermore, the formation

energies of interstitials, vacancies, and more complicated defects, are sufficiently different from one another that we can hope to identify the defects being removed in particular recovery stages.

Three stored energy experiments have been performed to date on irradiated copper, and one on aluminum. Of those done on copper, one covered recovery Stages II and III following deuteron irradiation (Overhauser, 1954), one Stage I following neutron irradiation (Blewitt, et al., 1956), and one Stage I following electron irradiation (Meechan and Sosin, 1959). Even though the experiments covered different recovery regions and used different kinds of irradiation, it might be expected that similar values would have been found for the ratio of resistivity recovery to stored energy release, since the lattice parameter, length, and resistivity measurements following deuteron irradiation indicate that Frenkel defect annihilation is the dominant recovery process in Stages I, II and III (sec. 2.2.). Contrary to expectations, the results are in rather sharp disagreement. This may be interpreted as indicating a different type of recovery in Stage I following electron and neutron irradiation and different mechanisms have been proposed (sec. 2.2.5.). However, this does not explain the lack of agreement between the deuteron and electron results. Consequently, it seems best to withhold judgment as to the reason for the lack of agreement until more experiments are performed.

It is likely that the energy release in Stage III depends on the temperature at which the specimen was irradiated, as Corbett, Smith, and Walker (1959) found for the recovery of resistivity. Therefore, it is important that the experiments covering Stages II and III be repeated and Stage I be studied, all following deuteron irradiation at temperatures in the liquid helium range. It also is important that the measurements following neutron and electron irradiation be extended upward in temperature through Stage III and all the way to complete recovery. Similar studies must be performed on other metals as well.

DeSorbo has recently measured the release of stored energy in quenched gold specimens. Additional experiments should be performed and the results compared with those obtained from irradiated specimens.

Since point defects can be trapped at impurities and dislocations, it should prove valuable to study the effects on the subsequent release of stored energy of both doping with various impurity atoms and of cold working prior to irradiation. Quenching of specimens from elevated temperatures prior to irradiation may also alter the manner in which the stored energy is released, since

additional vacancies will appreciably alter the relative concentration of interstitials and vacancies produced, particularly in the case of electron irradiation.

Greater experimental accuracy must be achieved before the above-mentioned kinds of experiments can be carried out. While the methods by which this accuracy will be achieved cannot be predicted, it might be suggested that more attention be given to the use of calorimetric measurements between well-defined states, i.e., between temperatures at which the recovery rate is negligible.

3.16. Calculations of the Interaction of Point Defects with Dislocations

The experimental observations of hardening processes in various alloys provoked a theoretical study of the effect of solute atoms and impurity agglomerates upon the behavior of dislocation lines (Cochardt, et al., 1955). Solute atoms act both to pin the dislocation lines and to impede their flow when the stress is sufficiently high to cause "breakaway." The behavior of point defects is very similar, but with the important difference that their number need not be conserved.

Internal friction studies have stimulated some calculations on the pinning of dislocation lines by point defects (Granato and Lücke, 1956). The effects of agglomerates of point defects on the flow of dislocations also have been considered (Seeger, 1958; Coulomb and Friedel, 1957). These considerations do not always take account of the detailed atomic nature of the point defects when they are associated with dislocations. There exist now various models to describe what happens when point defects arrive at a dislocation line:

(a) The defects retain their identity but have a certain binding energy with the dislocation, so that the dislocation becomes surrounded by an atmosphere of point defects.

(b) The point defects condense as agglomerates on the dislocation line.

(c) The point defects completely lose their identity by causing the dislocation line to climb. This is thought only to occur at jogs, but a special situation arises when there is a high supersaturation of interstitial atoms. This has not received consideration.

Theoretical work to decide which of the models is correct is greatly needed. Very likely, different models prevail under different circumstances, and there is now little guidance to assist in choosing the appropriate model. The interaction of moving defects,

including focused collision chains, with dislocation lines also may be important (Leibfried, 1959) and deserves further study.

The mobility of point defects along dislocation lines, although experimentally claimed, has not yet been treated properly from a theoretical standpoint. Current ideas are rooted in very naive considerations and much disagreement exists. In view of the importance of the problem, further calculations would be welcome.

The general problems associated with the vacancy climb of screw dislocations into spiral configurations and of edge dislocations into loops are becoming much clearer as a result of direct observations on quenched materials. Better models may now be applied to the interaction of vacancies with dislocations which also should pertain to the problems of the interactions of interstitials with dislocation lines. The resultant effects due to both vacancies and interstitials must be assessed in radiation problems.

3.17. More Use of Pedigreed Materials

Present data indicate that the most valuable experiments are those which are simplest and which involve high purity materials. If the materials investigated contain *any* impurities in concentrations approaching that of the defect being studied, difficulties can arise. The concentrations of displaced atoms after bombardment with 10^{17} particles per cm^2 are approximately as follows:

Radiation Concentration

1.4 Mev electrons	2.2×10^{-6}
10 Mev deuterons	5.5×10^{-4}
pile	1.2×10^{-5}

Experiments are desirable on materials whose behavior is well-known, both experimentally and theoretically. For example, quenching and irradiation studies on tungsten have the disadvantage that theoretical calculations concerning defect energies have not been made for tungsten. In this respect the alkali metals have been neglected. If they could be prepared in their low temperature phase, irradiation studies would be valuable. It is also possible that experiments on the noble gas solids might prove to be simple.

Aluminum recently has become available in a state of high purity; moreover we know a fair amount about its band structure, its elastic constants, etc. It would seem that it would serve as a good material of small atomic number with which to perform irradiation studies.

3.18. More Use of Doped and Pre-Treated Crystals

Corbett and Walker, and Blewitt and co-workers have thus far been the only research groups to deliberately introduce impurities and lattice defects into a specimen prior to low temperature irradiation. In both cases valuable information was obtained. Such experiments are only useful when both the nature of the irradiation damage and the nature of the added defect or impurity are reasonably well understood.

3.19. More Extensive Measurements of Threshold Energies for Displacement

A number of threshold energies for displacement have been determined. These have been summarized by Dienes and Vineyard (1957) (see also sec. 2.1.). The interpretation of the results of an experiment designed to measure a threshold displacement energy is usually not a simple, straightforward procedure. Some of the factors which make interpretation difficult are the following:

- (1) The bombarding electrons may not always be monoenergetic.
- (2) The finite thickness of the specimen degrades the energy of the bombarding electrons.
- (3) The probability of displacement of an atom is not a simple step-function of the energy transferred to it, as is frequently assumed in the interpretation of experimental results.
- (4) It may be possible to produce different kinds of defects in a given material, each with its own independent threshold displacement energy. Since each of these defects has its own recovery temperature, it is possible to observe apparent temperature-dependent threshold displacement energies.

The first two of these factors may be eliminated by the proper choice of experimental conditions. Efforts should be made in future experiments to use extremely thin specimens and monoenergetic electrons.

The third factor is perhaps the most difficult to treat. It is important to determine the probability of displacement, P , as a function of energy transfer, since the energy interval required for P to change from zero to unity is, in general, not small relative to the threshold energy. Most experimental measurements will indicate the threshold to be the energy at which P first differs by a measurable amount from zero. Obviously, if P is a slowly varying function of energy for small P , the experimentally determined threshold will decrease with increasing sensitivity of measure-

ment. The effective threshold displacement energy which should be used to calculate the number of primary displacements produced by more energetic irradiation, however, is not the minimum energy which has a non-zero probability of displacing an atom, but some intermediate energy in the energy interval over which P varies from zero to unity. If this interval is appreciable, the "effective threshold" will also vary with the kind of irradiation used, and we should perhaps not even speak of a threshold energy, but refer directly to the energy dependence of P . For any given kind of irradiation, then, instead of computing a cross section for displacement, $\sigma(\epsilon_D)$, on the assumption of a single threshold displacement energy, ϵ_D , we should use a differential displacement cross section, defined as

$$d\sigma_D = P(\epsilon) \frac{d\sigma}{d\epsilon} d\epsilon. \quad (3.2)$$

Here, $d\sigma/d\epsilon$ represents the derivative of the usual expression for the cross section, $\sigma(\epsilon)$, for transferring more energy than ϵ to an atom by a primary particle. The energy spectrum of the primary knock-ons is then given by

$$\frac{dn}{d\epsilon} = -N_0 \phi \frac{d\sigma_D}{d\epsilon}, \quad (3.3)$$

where ϕ is the integrated flux of primary particles, N_0 is the spatial density of atoms, and $n = n(\epsilon)$ is the number of primary knock-ons produced per unit volume with energies less than ϵ .

To determine $P(\epsilon)$ experimentally, it is necessary to choose a kind of irradiation for which $\sigma(\epsilon)$ is accurately known, and to monitor the change in a physical property to which the contribution of the damage produced by each primary knock-on is accurately known.

The effects of the fourth factor listed above should be studied by performing irradiation exposures at temperatures in the liquid helium range and by identifying all recovery stages produced by a relatively high energy electron irradiation. The dependence on electron energy of the rate of accumulation of the damage which recovers in each stage should then be investigated.

Not only is the threshold energy a fundamental parameter, determining, in the cascade theory, the number of atomic displacements produced by a knock-on, it may also be important for other reasons. If direct calculations of the displacement process are made (sec. 3.21.), it is found that the threshold depends sensitively on the interatomic force law in precisely the range where this force law is vital for radiation damage theory and yet is least

known. Accurate threshold measurements may thus allow interatomic force laws to be determined.

3.20. Measurements and Calculations of Atom-Atom Interaction Potentials

The basis for most of the differences in existing theories of radiation damage in metals lies in the choice of an atom-atom interaction potential. From a radiation damage point of view, the range of atomic separations over which it is important to know the strength of the interatomic potential is approximately 0.2 to 0.8 atomic diameters (1 to 4 a_0 for copper, where a_0 represents the radius of the first Bohr orbit for hydrogen, 0.528Å). A theoretical expression, given by Bohr (1940), has been used quite extensively. Its use, with screening constant as originally given, only can be justified at interatomic separations that are small relative to a_0 . Near one interatomic distance (approx. 5 a_0 for copper), we may expect the potential to be represented by a Born-Mayer expression, at least for close-packed metals, since the dominant interaction arises from the overlap of closed shells. The constants in the Born-Mayer expression can be evaluated from elastic constant data. When this is done for copper, the result is approximately four orders of magnitude larger than the extrapolated value of the Bohr expression near one interatomic distance. The Bohr expression, therefore, seems to seriously underestimate the strength of the atom-atom interaction in the range of importance for radiation damage calculations. Brinkman (1956b) has proposed another interaction potential. It somewhat overestimates the actual potential at atomic separations near an interatomic distance; however, it increases less rapidly with decreasing separation than do the Huntington potentials for copper (Huntington, 1953), and it crosses the extrapolation of the harder of the two Huntington expressions in the range of interest for radiation damage calculations. As of this time, it can only be stated that the Brinkman potential represents a guess as to the actual potential in the range of interest, and it cannot yet be concluded whether or not it is reasonable. A few experimental determinations of the atom-atom interaction potential over the range of interest should eliminate the wide discrepancy which now exists between various theoretical models of radiation damage in solids.

Perhaps the closest approach to the experimental determination of an interaction potential for metallic atoms is the study of knock-on ranges made by Schmitt and Sharp (1958), and the interpretation worked out by Holmes and Leibfried (1959). These

experiments do not determine the potential function, but they are able to show that the original Bohr expression is too short-ranged for use in radiation damage calculations.

Atom-atom interaction potentials for the inert gases can be obtained over fairly large ranges of interatomic separations by determining the high pressure equations of state. Present determinations cover the range from about one atomic diameter outward. While this does not cover the range of interest, it is possible to obtain some idea of both the magnitude and slope of the potential at the upper end of the range. Although the inert gases have not yet been made the subject of radiation effects experiments, the interatomic potential in the range of interest should depend primarily on the atomic number and be relatively insensitive to the chemical nature of the atoms. Furthermore, it may be possible to extend the range of the potential to somewhat smaller interatomic separations, e.g., 80 percent of an atomic diameter, by use of shock wave experiments to determine the equation of state.

It does not seem feasible to increase the pressure of a gas or solid to a level at which the equation of state can give information about the interatomic potential at separations much less than an atomic diameter. Therefore, to examine this range experimentally, it will be necessary to supply the necessary energy to the atoms by other than thermodynamic means. For such atoms as copper and silver, ion energies in the range 10 Kev to 100 Kev are required. The most reasonable experimental approach seems to be the study of the scattering of ions of such energies by neutral atoms. While this is not atom-atom scattering, it seems reasonable to assume that for reasonably heavy atoms the scattering of singly charged ions by neutral atoms will not differ very much from atom-atom scattering. In fact, it may be possible to measure the scattering of both singly and doubly charged ions and extrapolate to zero charge.

A little thought about such an experiment strongly indicates that it is important to achieve conditions wherein essentially only single scattering occurs. This rules out the possibility of passing the ion beam through any kind of solid foil. A layer of gas 1 cm thick at a pressure of 10^{-4} mm of Hg should be about right for copperlike atoms. This should singly scatter about 4 percent of the incident ions and doubly scatter only about 0.16 percent. For copper, this pressure corresponds to the vapor in equilibrium with liquid copper at a temperature of $1,200^{\circ}\text{C}$. It seems reasonable to maintain copper at such a temperature in an enclosure with entrance and exit holes provided for the ion beam.

Another possible method for learning about atom-atom interaction potentials is described briefly at the end of section 3.19. The problems of experimentally determining atom-atom interaction potentials are so severe that it will probably be several years before substantial results are available. It is important, therefore, that improved theoretical calculations of atomic interaction potentials be made. One Thomas-Fermi calculation has recently been published (Firsov, 1958), and others are in progress. It is, therefore, probable that better theoretical potentials soon will be available, but continued work on the problem, particularly in the direction of improving the Thomas-Fermi approximation, is highly desirable.

3.21. Direct Calculations Concerning the Displacement Process

The basic atomic event in radiation damage is initiated, as has been explained before, when an atom on a lattice suddenly becomes endowed with a rather large amount of energy, anywhere in the range from 10 to 100,000 ev. The theoretical ideas about the resulting processes have been formed mainly by considering in a qualitative way simplified models. To date, only two serious attempts to calculate a specific case in detail from first principles have been published. Both have been concerned with the question of the threshold energy for production of a stable defect (a close Frenkel pair, presumably). One calculation was made by Huntington (1954), for copper, the other by Kohn (1954), for germanium. The interatomic repulsive potential for intermediate distances of approach proves to be uncertain to a serious degree in such calculations (sec. 3.20.), but even after assuming a reasonable potential, many further assumptions must be made before the complicated many-body problem involved becomes tractable to ordinary analysis. Huntington and Kohn did find displacement thresholds which are not inconsistent with experimental results (the case of germanium remains somewhat controversial), but the estimated limits of error were wide. With the increasing complexity of experimental results, particularly with copper and the noble metals, and in view of the variety of models proposed in explanation, more powerful and more detailed calculations of the basic damage events are badly needed.

It now seems certain that numerical calculations of specific cases can be carried through with considerable accuracy on a modern high-speed computing machine. One technique is to program the machine to solve the equations of motion for the many-body system comprising the primary knock-on and all of its neigh-

bors that will be strongly affected. Initial conditions for the calculation can be that all atoms are on their lattice sites, the knock-on is endowed with a certain initial velocity, and all others are at rest. The machine can then compute a trajectory in configuration space and thus find the disturbed configuration of the set of atoms after all energy has been dissipated. With a large computer it seems possible to obtain fair accuracy in reasonable computing times with a set of the order of 500° to 1000 atoms. This is large enough to be realistic and to allow events of moderate energy (up to a few hundred eV) to be studied without serious disturbance from the boundaries of the set.

From work of this sort it may be possible to determine the probability of displacement as a function of energy in the threshold region, to assess the frequency with which the various kinds of close Frenkel pairs are formed, and to evaluate the importance of replacement collisions and focusing effects. It also may be possible to learn whether crowdions have any probability of being formed and their dynamic and static stability. Still another question of interest is the extent to which interstitials will tend to be formed in clusters, and whether divacancies and trivacancies will be produced at higher energies. Calculations of this type may also provide the theoretical insight needed to improve existing cascade and spike models or to decide upon the adequacy of either model.

3.22. Calculations of Energies Associated with Defects

Calculations of the energies associated with defects are extremely difficult and only a few of many calculations have survived. Both closed-shell repulsion and many electronic terms are important in the close-packed metals. At present we have reasonably good calculations for the energy required to form a lattice vacancy, an interstitial, and a crowdion in copper. It would be of value to calculate the energy of a di-interstitial in copper. It would also be valuable to decide, if possible, whether two kinds of interstitials can exist simultaneously in copper at low temperature, or whether the unstable type converts easily to the stable type. Moreover, it is not entirely certain whether the stable type is the point interstitial or the crowdion. Calculations on formation and motion energies for various defects in silver, gold, sodium, and aluminum would be valuable. Moreover, such calculations should be done for silicon and germanium.

3.23. Further Calculations of the Influence of Lattice Defects on Physical Properties

By and large, the effect of point defects upon the physical properties of copper have been calculated, but improved understanding awaits better knowledge of the atomic interaction potentials, among other things. For instance, the dilation, resistivity, formation energy, and mobilities of vacancies in copper have been calculated (Broom and Ham, 1957) and the results can be tested against experimental values for quenched materials, but the presence of divacancies, etc., make comparison uncertain. Knowledge of the influence of interstitials upon such properties is less well understood, largely because of the uncertain configuration of an interstitial atom and the inability experimentally to study this defect separately. Estimates of the influence of vacancies and interstitials on the physical properties of metals (other than copper) of differing crystallographic structure would assist greatly in the comparison and interpretation of experimental findings. Perhaps the most significant area for further calculations lies in the effects of point defects upon mechanical properties, i.e., in their interaction with dislocation lines (as discussed in section 3.16).

The effects of point defect clusters (both large and small), divacancies, trivacancies, etc., upon physical properties have not received sufficient attention, partly because the nature of the clusters themselves is little known. A knowledge of both the mobility and influence of these clusters is essential to the understanding of the annealing processes and the effects of radiation at high temperature. Since the larger clusters, e.g., those in quenched samples, can now be observed, their nature has been better defined, and calculations of their influence upon the physical properties are becoming relatively simpler.

3.24. Further Experiments on Growth and Swelling of Fissionable Materials

Many of the investigations of fissionable materials have been of a technological nature, aimed at answering specific questions quickly. Less effort has been placed on understanding the fundamental processes, as considered below.

The principal uncertainty about radiation-induced growth is whether it is due to a spike or a point-defect mechanism. A systematic study of the structure sensitivity and temperature dependence of the growth process would help place the theories on a more quantitative footing. Increasing numbers of point-defect

sinks can be introduced by controlled cold-work, and their effect in reducing the probability of Frenkel-pair recombination should be studied more fully. A better knowledge of the variation of growth rate with temperature should enable a more precise comparison with the rates predicted from the point-defect theories. The high temperature limit of growth would be dependent upon the flux, if the point-defect theories are correct. This could be tested easily by using uranium of various enrichments in the same neutron flux.

The radiation-induced creep of polycrystals that results from growth of individual grains might be used as a measure of growth rate in some of these studies. Polycrystalline springs, similar to those used by Roberts and Cottrell (1956), could provide large extension rates ($\dot{\epsilon}_e$), and the growth rate ($\dot{\epsilon}_g$), which by producing internal stresses causes the enhanced creep rates, could be deduced from the formula $\dot{\epsilon}_e = (\sigma/\sigma_y) \dot{\epsilon}_g$, in which σ is the stress applied to the spring and σ_y the yield stress.

The growth of other anisotropic materials doped with U^{235} or, more conveniently, with B or Li, might be simpler to handle experimentally and might produce less complex and more plentiful information upon which to test theories.

The lack of understanding of growth reflects ignorance of the fission fragment spikes. Further information could be gained by studying the homogenization of precipitates in alloys or the disordering of ordered alloys by the fission process. Alloys with precipitates of various sizes could be used to indicate the size of the spikes. A study of the precipitation of stable phases in metastable alloys or of the ordering of disordered alloys by fission fragments should indicate the importance of the point defects produced. If enhanced diffusion could be obtained, in an appropriate temperature range, then an estimate could be made of the number of point defects which survive the spike. A study of these opposing tendencies of point defect and spikes should permit an estimate of their relative importance, since the former promotes the establishment of equilibrium and the latter creates randomness and disorder.

Direct observation of the spikes will most likely lead to a clear understanding of the growth process. Their size is ideal for observation in the electron microscope, using foils about 1,000Å thick. It is difficult to anticipate the observations, but if the spike theories are to explain growth, then the nature of the spike should depend upon the crystallographic direction of the fission fragment. It may even be possible to estimate from the electron diffraction effects the number of atoms lost from the spike.

Swelling can be most simply studied by irradiation of specimens at temperatures too low for inert gas atoms to precipitate, followed by heat treatments to precipitate bubbles. The specimens should be examined metallographically on a sufficiently fine scale. These experiments have the advantage that any complications from the presence of neutron flux are avoided. They are capable of providing much information on the solubility and migration of gas atoms and on the nucleation and stability of the bubbles. Samples heated while in a neutron flux are less easily interpreted but also are essential to an understanding of the effect the neutron flux has upon these processes, e.g., the enhancement of diffusion rates of gaseous atoms within certain temperature ranges.

Detailed information on the nucleation and subsequent behavior of gas bubbles can be obtained by observation using an electron microscope. Replicas will prove useful for determining the number and location of bubbles, but thin film techniques are probably better suited to determining their size and their intersection with crystal defects, such as dislocation lines. The solubilities of the inert gases in various metals and alloys are not well known and should be studied since the amount of radiation a sample can sustain before forming bubbles at high temperatures is a function of the solubility limit. Alloys with "open" structures capable of accommodating the inert gases should be particularly investigated. The migration rate of the inert gases is likely to determine the ultimate size of the bubbles formed and can be investigated by observing diffusion of gases out of lightly irradiated samples.

It is important also to have information about the stability of such gas bubbles, for if they are able to coarsen, the volume increase becomes much greater. This stability depends upon the ability of the inert gas atom to re-enter into solution from the bubbles. Such coarsening has been observed in beryllium in which helium is produced by neutron bombardment (Rich, et al., 1959). The larger gaseous atoms, Xe and Kr, are generally believed to be unable to produce similar effects, but experimental confirmation is necessary. Observations of the size and number of bubbles in irradiated uranium after various heat treatments should be sufficient to demonstrate this.

The chemical activity and radioactivity of uranium alloys and compounds, with their complicated features of structure and anisotropy, combine to make detailed investigation difficult. Experiments using materials doped with elements which produce inert gases, e.g., lithium and boron, can be used for a detailed study, or the inert gases may themselves be injected into any material

by bombardment with fast ions of the gas. Thus, for example, alpha-particles can be used to inject helium.

3.25. More Quenching Experiments to Shed Light on Point Defects

Quenching experiments are of value since the "quenching-in" process produces a simpler defect pattern than irradiation. According to the calculations of Huntington and Seitz and of Tewordt, in the noble metals only vacancies and some vacancy clusters (divacancies, trivacancies, etc.) exist in equilibrium at high temperatures. Thus quenching studies can help distinguish the annealing associated with vacancies and vacancy clusters from that associated with interstitials and interstitial clusters.

To date the quenching data for gold and the quenching and equilibrium data for aluminum are the most complete. For gold, the present measurements are consistent with a vacancy mechanism for diffusion, with a vacancy formation energy $E_F^V = 0.98 \pm 0.03$ ev, and a vacancy motion energy $E_M^V = 0.82 \pm 0.05$ ev. The sum of these agrees well with the measured activation energy for self-diffusion in gold, i.e., $Q = 1.80 \pm 0.05$ ev.

For aluminum, Simmons and Balluffi (1958) prove that the defects present in equilibrium at high temperatures are lattice vacancies, since their length measurements indicate a larger volume per atom than their lattice parameter measurements. DeSorbo and Turnbull (1959) and Simmons and Balluffi both find $E_F^V = 0.79 \pm 0.06$ ev. In addition, the quenching experiments give $E_M^V = 0.65 \pm 0.06$ ev.

Preliminary quenching data exist for copper, silver and platinum, but the energies of formation and motion are probably not yet accurately known. Self-diffusion data, however, are available for these three metals: $Q_{Cu} = 2.05 \pm 0.05$ ev; $Q_{Ag} = 1.92 \pm 0.04$ ev; $Q_{Pt} = 2.96 \pm 0.06$ ev.

3.26. More Cold-Work Experiments to Shed Light on Defects

Experiments involving the cold-working of metals at 4°K followed by annealing are, in principle, more complex than either quenching or irradiation damage experiments, since dislocations, interstitial atoms, lattice vacancies, and various combinations of these defects are probably introduced by the former. Nevertheless, several points are worth mentioning.

Blewitt, Coltman, and Redman deformed copper single crystals at 4°K and Meehan and Sosin (1958) deformed polycrystalline copper, nickel, and gold at 4°K. Both groups observed

only minor annealing behavior on warming to 78°K (i.e., the drop in resistivity was from 1 to 3 percent of the increase resulting from cold-work). For deformed aluminum single crystals (Sosin and Koehler, 1952), the drop at 78°K was about 9 percent of the resistivity increment produced by deformation at 4°K.

It is clear that vacancies and interstitials should be produced by deformation. If so, and if a 35°K annealing stage in irradiated copper results from interstitial motion, then why is so little interstitial-vacancy recombination observed in the annealing of a cold-worked copper crystal? Why should vacancies and interstitials not behave similarly? It is possible that since interstitials have a large energy of formation, fewer interstitials than vacancies are formed. It is possible also that interstitial clustering produces a larger decrease in energy than does vacancy clustering; this is certainly true when identical size platelets of interstitials and vacancies are formed. Moreover, in cold-worked material it is likely that vacancies will be produced near other vacancies and interstitials near other interstitials; hence, clustering may be favored over interstitial-vacancy annihilation.

Buckel (1959) has carried out extensive measurements on the annealing of thin metal films prepared by evaporation at 4°K. The disorder in such films is large and shows complex annealing behavior. The similarity between this and the extended annealing stages seen after reactor irradiations has been noted and deserves further exploration.

3.27. Further Study of the Production of Color Centers in Alkali Halides

A major puzzle in the alkali halides is the failure of heavy particle irradiation to produce absorption spectra qualitatively different from those produced by X-rays. Interstitials and vacancies in both the cation and the anion system are expected from heavy particle irradiation. However, purely ionizing radiation should produce vacancies through the Seitz mechanism and interstitial anions through the Varley mechanism; but no interstitial cations should be produced. Thus, it must be concluded either that interstitial cations have no effect on the absorption spectra or that they are sufficiently mobile at all temperatures to migrate to sinks. The first possibility does not seem particularly attractive since the lattice strain associated with a cation would certainly give rise to localized exciton absorption. Moreover, in an interstitial site a cation should be an excellent electron trap which should be stable over some temperature range and, therefore, should intro-

duce an absorption band. The second possibility requires that an equal quantity of negatively-charged defects be removed at the same time as the cation interstitial to preserve electrical neutrality in the crystal.

A way out of this dilemma is the hypothesis that the same types of point defects, namely interstitials and vacancies in both the cation and anion lattices, are produced by both ionizing and fast particle radiation. Once again, however, this prospect is not too inviting for a number of reasons, the major one of which is the markedly different behavior of ionic conductivity after neutron bombardment and X- or gamma-ray exposure (Nelson, et al., 1958). Then, too, there is the problem of explaining such defect production by ionizing radiation. A solution of this problem would contribute greatly to the understanding of defects in ionic systems.

3.28. Elucidation of the Metamict State: Quartz

When vitreous silica is heavily irradiated in a reactor, a new amorphous state is produced, similar in many ways to the unirradiated material but of greater density and with several of its other properties appreciably changed. When quartz is heavily irradiated, its density decreases, and a state is reached which is identical, or nearly identical with that derived by irradiating vitreous silica (Wittels, 1957; Primak, 1959). This fully irradiated, or metamict state differs in rather subtle ways, in so far as atomic arrangement is concerned, from the vitreous state. Investigations by means of X-ray diffraction have indicated some of these differences (Simon, 1956, 1957), but have not yet fully characterized the metamict state. Also the mechanism by which irradiation produces this state is a matter of controversy. One view proposes that the large number of atoms displaced in cascade processes, particularly oxygen atoms that have lodged interstitially in positions outside the SiO_4 tetrahedra, and concomitant broken bonds are responsible. Another maintains that thermal spikes and their attendant strains are the cause. An increase in density with irradiation is an unusual phenomenon in itself,³ and the entire course of the transition in quartz, from the pristine to the fully irradiated state, is unusual, the apparent rate of transformation being slow at first, then accelerating, then finally becoming zero.

It would be most interesting to know whether similar new states can be produced in other materials. Evidently the directional

³ Boron glass is also known to increase in density under neutron irradiation, see Mylonas and Truell (1958).

bonding in silica, which permits it to exist in the vitreous state, also helps to stabilize its fully irradiated state, but such irradiated states have not been found in many other vitreous materials. The amorphous state caused by heavy irradiation of diamond (Levy and Kammerer, 1957) bears certain resemblances to that in silica, but has not been investigated as extensively. It is easy to imagine that ionization effects are also playing an important role, and it would be most interesting to irradiate both quartz and silica to very heavy exposures in an intense gamma ray field. Also it is noteworthy that all of the heavy irradiations of quartz and vitreous silica have been conducted at temperatures in the moderate temperature range of 30°C to 100°C. Prolonged exposure at very low temperatures or at much higher temperatures might reveal a different behavior. It is intriguing to speculate about the possible production of an amorphous state in every material if sufficient irradiation is given at a sufficiently low temperature. More extensive structural investigations of the irradiated state, using X-ray and neutron diffraction together, would also be useful in establishing details of atomic positions.

4. SPUTTERING

4.1. Introduction

A special case of radiation damage well-known to gas discharge physicists for more than a century is physical sputtering—the process of ejection of atoms from solids under ion bombardment. Physical sputtering is to be distinguished from chemical or reactive sputtering, which is a completely different process arising when certain reactive gas ions strike certain solids and form volatile compounds.

The subject of physical sputtering has been reviewed by Massey and Burhop (1952), and more recently by Wehner (1955). The extensive work done in this field indicates that many parameters are operative in physical sputtering. At the gas discharge side, the kinetic energy of the bombarding ion, its mass, and angle of incidence are important parameters. At the solid state side, the most important parameters are the mass of target atoms, the interatomic forces, crystal structure and orientation in the lattice, and finally the binding energy of atoms at the surface, i.e., the heat of sublimation. Experimental data on the influence of these and other less important parameters such as target temperature, surface condition, etc., have been accumulating rapidly since it became known that certain conditions need to be fulfilled in order

to obtain reliable and consistent results. Foremost among these conditions are that the ion energy and the angle of incidence should be well-defined, that back-diffusion of sputtered atoms to the target be prevented, and that sputtering overcome the formation of chemisorbed impurity layers at the target surface which may form continuously from background gases in the tube.

Two aspects of sputtering are of particular interest with respect to radiation damage problems. One is sputtering at very low ion energy (~ 100 ev), the other the atom ejection patterns observed when single crystals are sputtered.

At very low ion energy, events resulting from an impact are confined to only a small number of atoms near the place of impact. Measurements become difficult in this region because of the pronounced influence of target surface conditions and the extremely small sputtering yields. Unfortunately, atoms are sputtered from metals in the neutral state and this prevents their direct detection by sensitive electrical methods. These difficulties, however, are not of a principal nature and careful studies at very low ion energies promise to increase our knowledge of atomic collisions and interactions in an energy range which is the most interesting in radiation damage problems, thereby supplementing high-energy electron bombardment studies. Threshold energies for sputtering, i.e., the minimum energies necessary for sputtering to ensue, are probably linked in some way to displacement or replacement energies in solids.

The most interesting aspect of sputtering concerns the study of the directions in which atoms are sputtered preferentially from single crystal targets. It was found that atoms are ejected preferentially in closest neighbor directions, i.e., $[110]$ in fcc and $[111]$ in bcc and diamond lattice crystals. With increasing ion energy the second-nearest neighbor directions, i.e., $[100]$ in fcc and bcc crystals, come into play as additional ejection possibilities. This phenomenon results in interesting deposit patterns when atoms are sputtered from small (point source) single crystals. The closer study of these patterns may provide a very interesting tool in investigating the latest stages of a heavy particle radiation damage event, in particular the anisotropic propagation of energy from a collision center. Of much interest in this connection is a recent result by Thompson (1959) who found essentially the same phenomenon when he bombarded a gold single crystal foil from one side with 4.5 mev protons and obtained preferred ejection of gold atoms in closely packed directions at the other side.

4.2. Experimental Techniques

Most of the earlier work in sputtering was performed in the cold cathode glow discharge (cathode sputtering). Unfortunately, this simple, and therefore attractive, discharge functions well only at gas pressures where the mean path of ions and sputtered atoms is small compared to the tube dimensions. The bombarding ions then have an undetermined angle of incidence and a large energy spread and the transport of sputtered atoms away from the target is governed by diffusion laws, with many sputtered atoms being reflected to the target. Under these circumstances the tube geometry and gas pressure enter the picture and it is difficult to collect reliable and quantitative results with regard to the basic process.

These difficulties are eliminated at gas pressures where the mean free path of ions and atoms is larger than the tube dimensions. Two methods have been used successfully in these studies, one with ion beams and the other with low-pressure-supported plasmas. In the first method ions are extracted from a gas discharge through a constriction into a low pressure region (differential pumping) where they strike the target. It is important to select a discharge which operates at a low enough voltage so that only singly ionized atoms are produced or otherwise to provide mass and charge separation so that only one species is selected for bombardment. Ion beams have a basic disadvantage at low bombarding energies in that space charge effects limit the beam density to rather low values (order of $\mu\text{amp}/\text{cm}^2$). Under these conditions, unless the background of impurities in the system is extremely low, the target surface may become covered with an oxide or other layer faster than it can be sputtered away.

High ion current densities, even at low bombarding energies, can be achieved with the second method by immersing the target directly in a low pressure plasma of high density. The target then behaves essentially as a large negative Langmuir probe for which the bombarding conditions are easily controlled and well understood. The target becomes covered with a Langmuir sheath and ions bombard the target with an energy equivalent to the potential difference applied between the target and the plasma in the target vicinity. If the target is large compared to the thickness of the Langmuir sheath, the incidence of the ions becomes normal to the target surface. This second method allows control of the bombarding energy down to very small values, which is not possible in the two-electrode glow discharge, and space charge effects are minimized by confining them to the Langmuir sheath.

In mercury the plasma is conveniently created in a low voltage discharge of several amps current maintained between a Hg-pool cathode and an anode. In noble gases the pool cathode must be replaced by a hot cathode, e.g., an oxide cathode, or the plasma must be created by other means such as high frequency excitation. Properly designed magnetic fields allow substantial increases in plasma or ion current density, and an ion current density of 30 ma/cm² can be attained at gas pressures in the micron range. In yield determinations the number of ions which bombard the target must be known. The target current is not due exclusively to collected ions, however, but is due as well to those electrons released under positive ion bombardment (γ -coefficient). The γ -coefficient can be determined separately—below 1,000 ev it rarely exceeds 0.15 and seems to be constant on clean metal surfaces over a wide range of energies.

The number of sputtered atoms can be determined either by measuring the number of atoms lost from the target or the number collected by a receiver. Target loss can be determined by weight change, or the thickness of the removed layer can be measured after opening the demountable tube. More sensitive but less reliable methods, in which it is possible to monitor target loss to some degree without opening the tube, make use of the resistance change of target wires or layers which become thinner by sputtering or measure the thermionic emission change of composite targets from which surface layers are removed by sputtering (Th, Cs or W). In collection methods the weight increase of a receiver for sputtered material is measured either after opening the tube or by means of a built-in balance, or by optically monitoring the buildup of the sputtered deposit on a transparent collector. Relative yields can be determined by measuring the work function change of a probe when it becomes covered with a sputtered deposit (Wehner and Medicus, 1954) or by converting neutral sputtered atoms, such as those of the alkaline metals, by surface ionization on a hot platinum surface to ions which can then be measured as electrical currents (Bradley, 1954). Radioactive tracer methods provide possibilities for measuring very low yields at very low ion energies (Morgulis and Tischenko, 1956). Sputtered atoms may be converted as well into excited atoms which makes their mass spectrographic analysis possible. The difficulty with most collector methods is that the angular distribution of sputtered materials must be known. This distribution changes with the bombarding energy of the ions and must be determined separately.

Yield measurements at low ion energies have to be supplemented

by probe measurements in order to determine the exact value of the plasma potential in the target vicinity. Studies of deposit distributions or of ejection patterns are conveniently made by collecting the sputtered material on glass ribbons which can be bent to the desired form such that, for instance, the material is collected in one azimuth.

4.3. Established Results

1. Material is sputtered from metals mostly in the form of neutral atoms; from ionic compounds material may be ejected as ions.

2. Under oblique bombardment the atoms are ejected preferentially in directions away from the direction of incidence (Wehner, 1954).

3. Sputtering under oblique incidence is much more efficient in many cases than under normal incidence. An increase of one order of magnitude in yields has been observed, for example, at an iron target when the incidence of 400 ev ions was changed from normal incidence (90°) to oblique incidence of 30° (Wehner, 1959).

4. Atoms sputtered from metals have kinetic energies which are considerably higher than thermal evaporation energies. For instance, atoms sputtered from a Pt surface which is bombarded by normally incident Hg^+ ions of 400 ev energy were found to have average kinetic energies of 20 ev (Wehner, in press).

5. Sputtering yields (atoms/ions) are sensibly independent of the ion current density. If surface layer formation is not overcome by sputtering, the yield of many materials increases with ion current density (Wehner, 1957).

6. Some metal oxides have very low sputtering rates, e.g., oxides of Al, Zr, Ti, Th, Ta, and U.

7. The target temperature enters into the picture only if surface layers are involved or if the target material changes its crystal structure. In Hg most metal targets must be kept above 300°C in order to prevent the formation of a protective Hg film.

8. Sputtering yields of liquids such as Sn or Ga do not seem to be significantly different from those of similar solid metals.

9. Sputtering yields of alloys are not significantly different from those of their main constituents (Wehner, 1958a).

10. The angular distribution of material sputtered from polycrystalline targets by normally incident ions depends on the ion energy. At low ion energy the major part of sputtered atoms are ejected at low angles to the target surface. At energies in the range

of several kev, a cosine distribution is approached. At still higher energies, more atoms are ejected with full directional change of momentum and emerge normal to the target surface.

11. The etch effects observed microscopically on sputtered target surfaces reveal that differently oriented crystallites are attacked at somewhat different rates. Grain boundaries show up mostly as furrows, revealing a higher rate of attack (Wehner, 1958b).

12. Typical yield curves are shown by Wehner (1958a), always exhibiting a region in which yield rises proportionally with ion energy. It is possible to define a "cut-in" energy V_0 and a slope S and it has been observed that both values for one ion species vary systematically with the position of the target element in the periodic system. The experimentally determined values for Hg^+ -ion bombardment also have been reported by Wehner (1958a). The "cut-in" energies are not identical with true thresholds in that the yield curves at lower ion energy go over into a tail region where yields decrease more or less exponentially with decreasing ion energy. Experimental data in this tail region are rather sketchy and the question as to whether a well-defined threshold exists has not yet been settled. At higher ion energy, the yield curves flatten out and yields reach a maximum value at energies in the range of tens of kev. From that point on, yields tend to decrease with increasing ion energy.

13. Studies with metal single crystals have shown that atoms are ejected preferentially in closest-neighbor directions, e.g., [110] in fcc and [111] in bcc or diamond lattice crystals (Wehner, 1956). At higher ion energy, the second-nearest neighbor directions appear as additional ejection directions. Second-nearest neighbor ejection in the more closely packed fcc crystals requires higher bombarding energies than in bcc crystals. At higher bombarding energies, fcc (111) surfaces show an interesting change from a threefold symmetrical to a sixfold symmetrical pattern, just as if the lattice at the surface had changed from fcc to cph.

4.4. Theories of Sputtering

For years it was believed that physical sputtering was essentially an evaporation process. This concept first was formulated by von Hippel (1926) and later extended by Townes (1944). According to this theory the kinetic energy of an incident ion is transformed into heat at the point of collision. The temperature, as a function of time and distance from the impact, is calculated by standard methods of heat flow. Surface atoms are presumably

evaporated from the localized "hot spots" surrounding points of impact.

This theory was given strong support by an experiment by Seeliger and Sommermeyer (1955) who bombarded a silver or molten gallium target with A^+ ions of ~ 10 kev energy at various angles of incidence. In every instance they observed that the density of sputtered atoms was proportional to the cosine of the angle from the surface normal, as expected from an evaporation theory. However, there are many difficulties with such a theory. For instance, Compton and Langmuir (1930) noted that the localized "hot spot" should result in high thermionic emission, which has not yet been observed. This can be understood if it is realized that the lattice and the conduction electrons are so weakly coupled in a metal that the latter are not appreciably heated in sputtering events. We also might expect a strong temperature dependence of the sputtering rate, however, and this has not been observed. The applicability of macroscopic thermal quantities on a microscopic scale is open to question. The observed dependence of the sputtering rate on the incident angle and on the atomic number of atoms involved and a threshold energy for sputtering do not seem explicable by an evaporation theory. Other experiments, particularly the sputtering of single crystals, show directional effects which cannot be explained by an evaporation theory.

Another concept of the sputtering process, historically older than the evaporation theory, was suggested by Kingdon and Langmuir (1922, 1923) who visualized a momentum transfer process on the basis of their work on the sputtering of thorium from the surface of thoriated tungsten wires. One important result of this approach was the prediction of a threshold energy for sputtering, although the expression they obtained has been experimentally shown to be incorrect. These ideas were later extended by Lamar and Compton (1934), who calculated by a billiard ball model the rate of sputtering in addition to thresholds.

Recently, the momentum transfer concept has been exploited further in two theories considering the details of specific collision sequences involving a few atoms near the surface. One of these theories was proposed by Henschke (1957a, b), who assumed collisions between perfectly elastic spheres with radii determined by the largest closed electron shell. Energy is lost during the collisions by excitations of Debye waves. Only surface atoms may be sputtered, and these only if they receive energy greater than the heat of sublimation and momentum in an outward direction. The incident ion (if under normal incidence) collides with an atom,

or atoms, below the surface, rebounds, and may cause a surface atom to sputter. The main purpose of this model is to explain the single crystal spot patterns obtained by Wehner (1956) and threshold energies for sputtering. There are several difficulties with such a theory. Henschke's explanation for sputtering by incident ions more massive than target atoms seems questionable. In this case the incident ion supposedly strikes an atom below the target surface. This struck atom immediately strikes neighboring atoms in the direction of its motion. The initially incident ion again collides with these piled-up atoms whereupon it is reflected as from a solid wall and may sputter another lattice atom on its way out. This unrealistic view was objected to analytically by Langberg (1958) and Silsbee (1957).

The other theory of this type was proposed by Langberg (1958) who considers collision sequences involving knock-ons, assuming a Morse potential for the interatomic interactions. Morse potential constants are evaluated by compressibility data and the potential is extrapolated to very high interaction energies, where its accuracy becomes uncertain.

The one quantity Langberg calculates is the threshold energy for sputtering with Hg. It appears feasible to calculate threshold energies by theories of this type, but the calculation of sputtering rates would be a more difficult job.

The possibility of applying radiation damage theory to the problem of sputtering was first exploited by Keywell (1955). Keywell used hard sphere collisions which he concluded were applicable for ions at least as massive as He ions and with energy less than 6 kev. He derived a formula for the sputtering yield by use of elementary neutron cooling theory and the Seitz formula for the number of displacements produced by a recoil atom. The threshold energy for sputtering is considered to be the minimum energy that an incident ion must have to produce one displacement. The theory involves the displacement energy and an absorption coefficient (plus the probability of a penetration type collision for incident ions less massive than target atoms). The formula appeared useful in correlating some data in the kev range.

A somewhat different momentum transfer approach has been utilized by Harrison (1956, 1957) who applied the mathematical methods of neutron diffusion theory. He considered the interactions of four distribution functions describing the state of bombarding atoms and target atoms. These distribution functions are solutions of the transport equation, which was assumed valid for this application. This theory probably is not valid for incident particles of

energy less than several kev since binding energies are ignored. The sputtering yield can be determined by an integration over the velocity direction and energy. The yield depends on four parameters: (1) mass ratio, (2) trapping parameter, (3) mean free path ratio, and (4) threshold energy. The resulting formula for the yield is apparently very complex and was not given.

Another recent theory of sputtering at considerably higher energies was formulated by Goldman and Simon (1958). This theory is supposedly applicable when the range of incident particles is much greater than the average range of knock-ons. A formula was proposed for the sputtering yield.

4.5. Needs and Prospects

In summarizing, it is believed, for reasons discussed in more detail above, that the following aspects in sputtering may become the most interesting and rewarding ones with relation to the field of radiation effects and need further study:

1) Sputtering at very low ion energies—reliable experimental data on yields and thresholds should be collected for different ion-metal combinations and attempts should be made to establish correlations with radiation damage parameters such as displacement energies and interatomic forces.

2) Ejection patterns obtained in single crystal sputtering—the closer study of these patterns for different ion-metal combinations at different bombarding energies is a new and unexplored field which promises to develop into an interesting tool for displaying directly the anisotropic energy transfer at the last stage of a radiation event in a crystal lattice.

5. REFERENCES

- Adams, M. A., and Higgins, P. R. B., 1959, *Phil. Mag.*, **4**, 77.
Antal, J. J., Weiss, R. J., and Dienes, G. J., 1955, *Phys. Rev.*, **99**, 1081.
———, and Goland, A. N., 1958, *Phys. Rev.*, **3**, 1280.
Aronin, L. R., 1954, *J. Appl. Phys.*, **25**, 344.
Atkinson, H. H., 1958, *Phil. Mag.*, **3**, 476.
Barner, R. S., 1952, *Proc. Phys. Soc.*, **80**, 207.
———, Churchman, A. T., Curtis, G. C., Eldred, V. W., Enderby, J. A., Foreman, A. J., 1958, *Second Int. Conf. on the Peaceful Uses of Atomic Energy*, Geneva, Paper No. 81.
———, and Greenwood, G. B., 1958, *Second Int. Conf. on the Peaceful Uses of Atomic Energy*, Geneva, Paper No. 29.
———, Redding, G. B., and Cottrell, A. H., 1958, *Phil. Mag.*, **3**, 97.
Baroody, E. M., 1958, *Phys. Rev.*, **112**, 1571.
Bauerle, J. E., and Koehler, J. S., 1957, *Phys. Rev.*, **107**, 1493.
Bemski, G., 1959, *J. Appl. Phys.*, **30**, 1195.

- , and Augustyniak, W. M., 1957, *Phys. Rev.*, **108**, 645.
- Berman, R., 1951, *Proc. Roy. Soc.*, **A208**, 90.
- , Foster, E. L., and Rosenberg, H. M., 1955, *Rept. of Bristol Conf. on Defects in Crystalline Solids* (Physical Society, London), p. 321.
- Binder, D., and Sturm, W. J., 1955, *Phys. Rev.*, **99**, 603; 1957, *ibid.*, **107**, 106.
- Bleiberg, M. L., Jones, L. J., and Lustman, B., 1956, *J. Appl. Phys.*, **27**, 1270.
- Blewitt, T. H., and Coltman, R. R., 1954, *Acta Met.*, **2**, 549.
- , Coltman, R. R., Noggle, T. S., and Holmes, D. K., 1956, *Bull. Am. Phys. Soc.*, Ser. II, **1**, 130.
- , Coltman, R. R., Klabunde, C. E., and Noggle, T. S., 1957, *J. Appl. Phys.*, **28**, 639.
- , Coltman, R. R., and Klabunde, C. E., 1959, *Phys. Rev. Letters*, **3**, 132.
- Bloch, R., 1953, *Sitzb. Osterr. Ak. Wiss. Math-naturw. Kl.*, **163**, 99.
- Blount, R. I., 1959, *Phys. Rev.*, **113**, 995.
- Bohr, N., 1940, *Kgl. Danske Videnskab Selskab, Mat.-fys. Medd.*, Vol. XVIII, No. 8.
- Bradley, R. C., 1954, *Phys. Rev.*, **93**, 719.
- Brinkman, J. A., 1954, *J. Appl. Phys.*, **25**, 961.
- , 1956a, *Am. J. Phys.*, **24**, 246; 1956b, *J. Appl. Phys.*, **25**, 246.
- , Dixon, C. E., and Meechan, C. J., 1954, *Acta Met.*, **2**, 38.
- Broom, T., and Ham, R. K., 1957, *Symposium on Vacancies and Other Point Defects in Alloys*, Institute of Metals, London.
- Brown, W. L., and Augustyniak, W. M., 1959, *J. Appl. Phys.*, **30**, 1300.
- , Augustyniak, W. M., and Waite, T. R., 1959, *J. Appl. Phys.*, **30**, 1258.
- , Fletcher, R. C., and Wright, K. A., 1954, *Phys. Rev.*, **96**, 834.
- Buckel, W., 1959, *Z. f. Physik*, **154**, 474.
- Chang, R., 1957, *J. Appl. Phys.*, **28**, 385.
- Clark, C. D., Ditchburn, R. W., and Dyer, H. B., 1956, *Proc. Roy. Soc.*, **A224**, 363.
- Cleland, J. W., Crawford, J. H., Jr., and Pigg, J. C., 1955, *Phys. Rev.*, **98**, 1742.
- , Crawford, J. H., Jr., and Holmes, D. K., 1956, *Phys. Rev.*, **102**, 722.
- Cochardt, A. W., Schoek, G., and Wiedersich, H., 1955, *Acta Met.*, **3**, 533.
- Compton, K. T., and Langmuir, I., 1930, *Rev. Mod. Phys.*, **2**, 123 or 186.
- Cooper, H. G., Koehler, J. S., and Marx, J. W., 1955, *Phys. Rev.*, **97**, 599.
- Corbett, J. W., Smith, R. B., and Walker, R. M., 1959, *Phys. Rev.*, **114**, 1452, 1460.
- Coulomb, P., and Friedel, J., 1957, in *Dislocations and Mechanical Properties of Crystals*, John Wiley and Son, New York.
- Coulson, R. A., and Kearsley, M. J., 1957, *Proc. Roy. Soc.*, **A241**, 433.
- Crawford, J. H., Jr., and Wittels, M. C., 1958, *Proc. Second Int. Conf. on the Peaceful Uses of Atomic Energy*, Geneva, Paper No. 679.
- , 1958, *Proc. of Int. Conf. on Solid State Physics*, Brussels (to be published).
- , and Cleland, J. W., 1959, *J. Appl. Phys.*, **30**, 1204.
- Damask, A. C., 1958, *J. Appl. Phys.*, **29**, 1590.
- , Dienes, G. J., and Weizer, V. G., 1959, *Phys. Rev.*, **113**, 781.
- DeSorbo, W., and Turnbull, D., 1959, *Acta Met.*, **7**, 83.
- Dieckamp, H., and Sosin, A., 1956, *J. Appl. Phys.*, **27**, 1416.
- Dienes, G. J., 1952, *Phys. Rev.*, **86**, 228; **87**, 666.
- , and Damask, A. C., 1958, *J. Appl. Phys.*, **29**, 1713; *J. Phys. Chem. Solids*, in press.

- , and Gurinsky, D., 1957, Proc. French-American Conf. on Graphite Reactors (Brookhaven National Laboratory), pp. 145-156.
- , and Vineyard, G. H., 1957, *Radiation Effects in Solids* (Interscience Publishers, New York).
- Dugdale, R. A., 1956, Phil. Mag., Ser. 8, 1, 537.
- Ells, C. B., and Perryman, E. C. W., 1959, J. Nuc. Materials, 1, 73.
- Fehér, G., 1937, Phys. Rev., 105, 750.
- Firsov, O. B., 1958, J. Exptl. Theoret. Phys. (U.S.S.R.), 34, 447.
- Fisher, J. C., Thomson, R., and Vreeland, J. A. (eds.), 1956, *Dislocations and Mechanical Properties of Crystals*, Wiley, New York.
- Footé, F. C., 1955, Proc. First Int. Conf. on Peaceful Uses of Atomic Energy, Geneva, Paper No. 555.
- Fujita, F. E., and Gonser, U., 1958, J. Phys. Soc. Japan, 13, 1068.
- Gilman, J. J., and Johnston, W. G., 1956, J. Appl. Phys., 27, 1018.
- , and ———, 1958, J. Appl. Phys., 29, 877.
- Gobeli, G. W., 1958, Phys. Rev., 112, 732.
- Goldman, D. T., and Simon, A., 1958, Phys. Rev., 111, 883.
- Gonser, U., and Okkerse, R., 1957, Phys. Rev., 105, 757.
- Gossick, R. R., and Crawford, J. H., Jr., 1957, Bull. Am. Phys. Soc., 3, 400.
- , 1959, J. Appl. Phys., 30, 1214.
- Granato, A., and Lücke, K., 1956, J. Appl. Phys., 27, 583, 789.
- Griiffiths, J. H. E., Owen, J., and Ward, I. M., 1955, *Rept. of Conf. on Defects in Crystalline Solids* (Phys. Soc., London), p. 81.
- Harris, E. G., 1955, Phys. Rev., 98, 1151. Naval Research Laboratory Report NRL-4807, Aug. 3, 1956 (unpublished).
- Harrison, D., Jr., 1956, Phys. Rev., 102, 1473; 1957, *ibid.*, 105, 1202.
- Harwood, J. J., Hausner, H., Morse, J. G., and Rauch, W. G., 1958, *Effects of Radiation on Materials* (New York, Reinhold).
- Henning, G., Dienes, G. J., and Kosiba, W., 1958, Second Int. Conf. on Peaceful Uses of Atomic Energy, Geneva, Paper No. 1778.
- Henschke, E. B., 1957a, J. Appl. Phys., 28, 411; 1957b, Phys. Rev., 106, 737.
- Hirsch, P., Silcox, J., Smallman, R. E., and Westmacott, K. H., 1958, Phil. Mag., 3, 897.
- Holmes, D., and Leibfried, G., 1959, J. Appl. Phys. (in press).
- Hon, J. F., and Bray, P. J., 1959, U.S.A.E.C. Report, Brown University.
- Huntington, H. B., 1953, Phys. Rev., 91, 1092.
- , 1954, Phys. Rev., 93, 1414.
- James, H. M., and Lark-Horovitz, K., 1951, Z. Physik Chemie, 198, 107.
- Johnston, W. G., and Gilman, J. J., 1959, J. Appl. Phys., 30, 129.
- Keywell, F., 1955, Phys. Rev., 97, 1611.
- Kinchin, G. H., and Pease, R. S., 1955, "The Displacement of Atoms in Solids by Radiation," Rept. Progr. in Phys., 18, 1.
- Kingdon, K. H., and Langmuir, I., 1922, Phys. Rev., 20, 108; 1923, *ibid.*, 22, 148.
- Klontz, E. E., and Lark-Horovitz, K., 1952, Phys. Rev., 86, 643.
- Kobayashi, K., 1956, Phys. Rev., 102, 848; 1957, *ibid.*, 107, 41.
- Kohn, W., 1954, Phys. Rev., 94, 1409.
- Konobeevsky, S. T., 1956, J. Nuc. Eng., 1, 356.
- Kosiba, W. L., and Dienes, G. J., 1957, *Advances in Catalysis*, Vol. IX, 398.
- Lamar, E. S., and Compton, K. T., 1934, Science, 80, 541.
- Lambert, M., and Guinier, A., 1957a, Compt. Rend, 244, 2791; 1957b, *ibid.*, 245, 526; 1958, *ibid.*, 246, 1678.

- Langberg, E., 1958, *Phys. Rev.*, **111**, 91.
Leibfried, G., 1959, *J. Appl. Phys.*, **30**, 1388.
Levy, P. W., and Kammerer, O. F., 1957, *Phys. Rev.*, **100**, 1787.
Lillie, D. W., 1958, *Proc. Second Int. Conf. on Peaceful Uses of Atomic Energy*, Geneva, Paper No. 186X.
Lomer, W. M., 1954, A.E.R.E. Report T/R 1540.
MacKay, J. W., Klontz, E. E., and Gobeli, G. W., 1959, *Phys. Rev. Letters*, **2**, 146.
Magnuson, G. D., Palmer, W., and Koehler, J. S., 1958, *Phys. Rev.*, **109**, 1990.
Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena*, Oxford Univ. Press, New York.
Maurin, J., Ballantine, D., and Sucher, J., 1959, *J. Chem. Phys.* (in press).
Mayer, G., Perio, P., Gigon, J., and Tournarie, M., 1956, *Proc. Int. Conf. on Peaceful Uses of Atomic Energy*, United Nations, Vol. 7, p. 647.
Meehan, C. J., and Sosin, A., 1959, *Phys. Rev.*, **113**, 422.
———, and Sosin, A., 1958, *J. Appl. Phys.*, **29**, 738.
Menter, J. W., 1958, *Phil. Mag.*, Supplement, **7**, 299.
Mitchell, E. W. J., and Wedepohl, P. T., 1958, *Phil. Mag.*, **3**, 1280.
———, and Paige, E. G. S., 1958, *Phil. Mag.*, **1**, 1085.
Morgulis, N. D., and Tischenko, V. D., 1956, *J. Exptl. and Theor. Phys.* (U.S.S.R.), **30**, 52.
Muller, F. W., 1957, *J. Appl. Phys.*, **28**, 1.
Mylonas, C., and Truell, R., 1958, *J. Appl. Phys.*, **29**, 1252.
Nabarro, F. R. N., 1952, *Phys. Rev.*, **87**, 665.
Nelson, C. M., Sproull, R. L., and Caswell, R. S., 1953, *Phys. Rev.*, **90**, 364.
Noggle, T. S., and Stiegler, J. O., 1959, *J. Appl. Phys.*, **30**, 1279.
Overhauser, A. W., 1954, *Phys. Rev.*, **94**, 1551.
Owen, O. S., and Holmes, D. K., 1959, *J. Appl. Phys.*, **30**, 1289.
Paine, S. H., and Kittel, J. H., 1955, *Proc. First Int. Conf. on Peaceful Uses of Atomic Energy*, Paper No. 745.
Palmer, W., and Koehler, J. S., 1958, *Bull. Am. Phys. Soc.*, **3**, No. 7, 366.
Paneth, H., 1950, *Phys. Rev.*, **80**, 708.
Pease, R. S., 1954, *Acta Cryst. Camb.*, **7**, 633.
Petch, N. J., 1953, *J. Iron and Steel Inst.*, **174**, 25.
Primak, W., *J. Phys. Chem. Solids* (to appear).
Pugh, S. F., 1955, *Proc. First Int. Conf. on Peaceful Uses of Atomic Energy*, Geneva, Paper No. 443.
Rappaport, P., and Loferski, J. J., 1955, *Phys. Rev.*, **100**, 1261.
Rich, J. B., Redding, G. B., and Barnes, R. S., 1959, *J. Nuc. Materials*, **1**, 96.
Rimmer, D. E., and Cottrell, A. H., 1957, *Phil. Mag.*, **2**, 1345.
Ring, P. J., O'Keefe, J. G., and Bray, P. J., 1958, *Phys. Rev. Letters*, **1**, 453.
Roberts, A. C., and Cottrell, A. H., 1956, *Phil. Mag.*, **1**, 711.
Ruby, S. L., Schupp, F. D., and Wolley, E. D., 1958, *Phys. Rev.*, **111**, 1493.
Schmitt, R. A., and Sharp, R. A., 1958, *Phys. Rev. Letters*, **1**, 445.
Schulz-Dubois, E., Nisenoff, M., Fan, H. Y., and Lark-Horovitz, K., 1955, *Phys. Rev.*, **98**, 1561.
Seeger, A., 1958, *Second Int. Conf. on Peaceful Uses of Atomic Energy*, Geneva, Paper No. 998.
Seeliger, R., and Sommermeyer, K., 1955, *Z. Physik*, **93**, 692.
Seitz, F., 1954, *Rev. Mod. Phys.*, **26**, 7.

- , and Koehler, J. S., 1956, "Displacement of Atoms During Irradiation," in *Solid State Physics* (editors, F. Seitz and D. Turnbull, Academic Press, New York), Vol. 2, pp. 307-442.
- Siegel, S., 1949, *Phys. Rev.*, **75**, 1823.
- Silsbee, R. H., 1957, *J. Appl. Phys.*, **28**, 1246.
- Simmons, R. O., 1959, *Phys. Rev.*, **113**, 70.
- , and Balluffi, R. W., 1958, *Phys. Rev.*, **109**, 885.
- Simnad, M. T., Smoluchowski, R., and Spilners, A., 1958, *J. Appl. Phys.*, **29**, 1630.
- , Smoluchowski, R., 1955a, *Phys. Rev.*, **99**, 1891; 1955b, *J. Chem. Phys.*, **23**, 1961.
- Simon, I., 1956, *Phys. Rev.*, **103**, 1587; 1957, *J. Am. Ceram. Soc.*, **40**, 150.
- Smallman, R. E., and Willis, B. T. M., 1957, *Phil. Mag.*, **2**, 1018.
- Smoluchowski, R., 1956, *Proc. Int. Conf. on Peaceful Uses of Atomic Energy*, United Nations, Paper No. 748.
- Sosin, A., Brinkman, J. A., and Meechan, C. J., 1959, *Physics Today*, **12**, 6, 30.
See also Report of Conference on Lattice Defects in Noble Metals, 1959, Oct. 16-17, Canoga Park, Calif., U.S.A.E.C. Report NAA-SR-3250.
- , and Koehler, J. S., 1956, *Phys. Rev.*, **101**, 972.
- Spaepen, J., 1958, *Phys. Rev. Letters*, **1**, 453.
- Stevens, D. K., Sturm, W. J., and Silsbee, R. H., 1958, *J. Appl. Phys.*, **29**, 66.
- Taylor, E. H., and Kohn, H. W., 1957, *J. Am. Chem. Soc.*, **79**, 252.
- Tewordt, L., 1958, *Phys. Rev.*, **109**, 61.
- Thomas, D. E., 1956, *Nuclear Metallurgy*, Am. Inst. Mining and Metallurgical Engineers, **3**, 13.
- Thompson, M. V., 1959, *Phil. Mag.*, **4**, 139.
- Thompson, D. O., and Holmes, D. K., 1956, *J. Appl. Phys.*, **27**, 713.
- , and ———, 1957, *J. Phys. Chem. Solids*, **1**, 275.
- Townes, C. H., 1944, *Phys. Rev.*, **65**, 319.
- Truett, R., Teutonico, L., and Levy, P. W., 1957, *Phys. Rev.*, **105**, 1723.
- von Hippel, A., 1926, *Ann. Physik*, **81**, 1043.
- Vook, F. L., and Balluffi, R. W., 1959a, *Phys. Rev.*, **113**, 62; 1959b, *Phys. Rev.*, **113**, 72.
- Vook, R. W., and Wert, C. A., 1958, *Phys. Rev.*, **109**, 1529.
- Watkins, G. D., Corbett, J. W., and Walker, R. M., 1959, *J. Appl. Phys.*, **30**, 1198.
- Wechsler, N. F., and Kernohan, R. H., 1958, *J. Phys. Chem. Solids*, **7**, 307.
- Weeks, R. A., 1956a, *J. Appl. Phys.*, **27**, 1376; 1956b, *Bull. Am. Phys. Soc.*, **3**, 136.
- Wehner, G., 1954, *J. Appl. Phys.*, **25**, 270.
- , 1955, *Advances in Electronics and Electron Physics*, Vol. VII, Academic Press, New York.
- , 1956, *Phys. Rev.*, **102**, 690.
- , 1957, *Phys. Rev.*, **103**, 35.
- , 1958a, *Phys. Rev.*, **112**, 1120; 1958b, *J. Appl. Phys.*, **29**, 217.
- , 1959, *J. Appl. Phys.*, **30**, 1762.
- , and Medicus, G., 1954, *J. Appl. Phys.*, **25**, 698.
- Wertheim, G. K., 1957, *Phys. Rev.*, **105**, 1730; 1958, *ibid.*, **110**, 1272.
- , and Buchanan, D. N. E., 1959, *J. Appl. Phys.*, **30**, 1232.
- Wertz, J. E., Auzins, P., Weeks, R. A., and Silsbee, R. H., 1957, *Phys. Rev.*, **107**, 1535.
- Westrum, E. F., 1956, *Proc. of Fourth Int. Conf. on Glass*, Paris.

- Whapham, A. D., 1958, *Phil. Mag.*, *3*, 103.
Wieninger, L., and Adler, N., 1956, *Acta Physica Australica*, *4*, 81.
Wilsdorf, H. G. E., 1959, *Phys. Rev. Letters*, *3*, 172.
Wittels, M. C., 1957, *Phil. Mag.*, *3*, 1445.
Young, D. A., 1958, *Nature*, *182*, 875.
Young, F. W., 1959, *Bull. Am. Phys. Soc.*, *II*, *4*, 136.
Zener, C., 1949, *Acta Cryst.*, *2*, 163.

Part XII

TECHNIQUES AND INSTRUMENTATION

Chairman

Prof. POL DUWEZ
California Institute of Technology
Pasadena, Calif.

Members

Prof. C. S. BARRETT
University of Chicago
Chicago, Ill.

Dr. R. S. DAVIS
A. D. Little, Inc.
Cambridge, Mass.

Prof. H. TRACY HALL
Brigham Young University
Provo, Utah

Dr. BENJAMIN LAX
Lincoln Laboratories
Massachusetts Institute of
Technology
Lexington, Mass.

Dr. O. C. SIMPSON
Argonne National Laboratory
Lemont, Ill.

Prof. ROHN TRUELL
Brown University
Providence, R. I.

CONTENTS

	Page
1. INTRODUCTION AND ACKNOWLEDGMENTS.....	715
2. MICROTECHNIQUES AND GENERAL.....	716
2.1. Observations of Dislocations.....	716
2.1.1. Optical and Etch Pit Methods.....	716
2.1.2. X-Ray Methods.....	718
2.1.3. Electron Microscope Methods.....	719
2.1.4. Point Emission Microscope Methods.....	721
2.2. Microradiography and Autoradiography.....	721
2.3. Small-Angle X-Ray Scattering.....	722
2.4. Diffuse Scattering.....	723
2.5. Precision Structure Determinations.....	724
2.6. X-Ray Spectroscopy.....	725
2.7. High Intensity X-Ray Sources.....	727
2.8. Electron Probe X-Ray Microanalysis.....	727
2.9. References.....	728
2.10. Reviews.....	729

3. HIGH PRESSURE, HIGH TEMPERATURE	780
3.1. Static Pressures and Temperatures Obtainable Today	780
3.2. Pressure Media	780
3.3. Conventional Apparatus	781
3.4. Massive Support Principle	782
3.5. Pseudo-Multistaging	783
3.6. Belt Apparatus	783
3.7. Possibilities of Higher Pressures for the Future	783
3.8. Effects to be Expected at High Pressure	785
3.9. Melting Phenomena	786
3.10. Synthesis of New Materials	786
3.11. Geology	787
3.12. Metallurgical Applications	787
3.13. Conclusions	787
3.14. References	788
4. HIGH TEMPERATURE	788
4.1. Electrical Resistance and Induction Heating	788
4.2. Solar Furnaces and Refocused Electric Arcs	740
4.3. Chemical Flames	741
4.4. Electric Arcs and Plasma Jets	741
4.5. Electron Beams	742
4.6. High Temperature and New Materials	743
4.7. References	744
5. HIGH INTENSITY STRESS WAVES	744
5.1. Introduction	744
5.2. The Rod Configuration	745
5.3. The Plate Configuration	745
5.4. The Source of High Pressures	746
5.5. Measurements of the Shock Wave Parameters	746
5.6. Structure and Property Changes	749
5.7. References	750
6. HIGH MAGNETIC FIELDS	751
6.1. Introduction	751
6.2. Experimental Programs	752
6.2.1. Cyclotron Resonance in Solids	752
6.2.2. Magneto-optic Effects	752
6.2.3. Galvanomagnetic and Thermomagnetic Effects	752
6.2.4. Magnetic Resonance	752
6.2.5. De Haas-Van Alphen Effect	753
6.2.6. Low Temperature Research	753
6.2.7. Magneto-Acoustic Resonance and Oscillatory Effects	753
6.2.8. Millimeter and Submillimeter Wave Generation	753
6.2.9. Mechanical Properties of Solids	753
6.3. Present Possibilities	754
6.4. Future Possibilities	754
6.4.1. Large Water-Cooled Coils	754
6.4.2. Transient or Intermittent Operation (Flywheels)	754
6.4.3. Low Temperature Operation	755
6.5. Pulse Magnets	755
6.6. Superconducting Magnets	756
6.7. References	757
6.8. Review Articles	757
7. HIGH FREQUENCY ULTRASONICS	758
7.1. Introduction	758
7.2. Attenuation Mechanisms	760
7.2.1. Dislocation Damping	760
7.2.2. Thermoelastic Effects (Umklapp Processes)	760

7.2.3. Phonon-Phonon Interactions	761
7.2.4. Scattering Processes	762
7.2.5. Electron-phonon Interaction (Metals)	762
7.3. Equipment and Techniques	763
7.4. References	765
8. ANALYTICAL TECHNIQUES	766
8.1. Introduction	766
8.2. Chemical and Spectrographic Methods for Small Concentrations	766
8.3. Physical Methods for Small Concentration Analysis	767
8.4. References	771

TECHNIQUES AND INSTRUMENTATION

1. INTRODUCTION AND ACKNOWLEDGMENTS

Instrumentation and techniques have always played an important part in the advancement of science and technology. This statement is particularly true when applied to the science of materials. Many new theoretical ideas were firmly established only after experimental evidence had been found through some new or improved techniques. Thus, the theory of dislocations in crystals was generally accepted as an interesting concept leading to a satisfactory explanation of plastic deformation. Improved techniques in metallography lead to actual visualization of dislocation etch pits. Rather recent developments in the technique of transmission electron microscopy resulted in a spectacular confirmation of the existence of dislocations. By this technique, the motion of dislocations under the influence of heat and stress could be studied in detail and as a result the foundations on which the general theory of dislocations is built, have been solidly confirmed on the basis of experimentally observed facts. The fundamental ideas concerning the energy band structure in semiconductors, were also confirmed by experimental studies following the development of a new technique, namely, cyclotron resonance.

The incentive for developing new techniques often comes from the need for confirming or disproving theoretical concepts. New techniques may also be motivated by a desire for improving the accuracy of measurements of physical phenomena. In any case, improved techniques, as well as really new ones, have always resulted in advances (some modest and others spectacular) in the understanding of the theory of solids.

In preparing this section of the report the panel tried to emphasize new techniques. This objective proved to be extremely difficult for the simple reason that new techniques are the result of inventions and specific inventions are unpredictable. Most of the mate-

rial presented will therefore deal with foreseeable improvements of already known experimental techniques, and the ways in which these improvements could lead to a better understanding of materials.

This panel report makes no attempt to be comprehensive or exhaustive; a limited number of techniques were selected for discussion. In the opinion of the authors, improvements in these techniques during the next few years, could bring about new perspectives in the field of materials.

Among the techniques which are not included in this section of the report, the following appear to be important enough to warrant brief mention: thermal analysis, including high precision calorimetry; optical techniques such as multiple beam interferometry; electron and ion field emission microscopy; ultra-high vacuum; precision density measurements; internal friction and relaxation; high sensitivity magnetic resonance; irradiation effects as a tool. Some of these techniques are discussed as part of the general text of other chapters of this book.

In the course of preparing this chapter, the panel has benefited greatly from the advice and criticism of: R. W. Bane, Miss M. A. Barrett, H. E. Bömmel, F. R. Boyd, Harvey Brooks, J. E. Dorn, A. M. Gaudin, Lawrence Himmel, L. P. Hunter, E. Jacobson, H. H. Kolm, A. C. Lloyd, J. B. Newkirk, M. E. Nicholson, E. R. Parker, K. S. Singwi, D. K. Stevens, D. T. Stevenson, F. S. Tomkins, C. Walker, J. M. Walsh, W. W. Webb, W. B. Wilson, D. B. Wittry and D. S. Wood.

2. MICROTECHNIQUES AND GENERAL

2.1. Observations of Dislocations

In recent years great advances have been made in observing individual dislocations and other inhomogeneities in crystalline materials (for recent reviews with appropriate literature references see section 2.10 below, and also Part VII on Growth, Structure, and Morphology of Crystals, and Part IX on the Nature and Properties of Solid Surfaces and Interfaces).

2.1.1. Optical and Etch Pit Methods

The contour of the surface of a crystal after growth under suitable conditions may show the points of emergence of screw dislocations, the step heights may disclose the magnitude of the Burgers vector, and the direction of winding of a spiral may indicate the sign of the dislocation. The technique is obviously limited in

its possible applications; a more fruitful method has been to employ etching solutions that develop pits at the points of emergency of dislocations. Although the etching characteristics depend upon the impurities present in the crystal and on whether these are segregated in the dislocations or not, on the etching solution, and on the method of preparation of the surface, it is frequently possible to control conditions so that reliable observations of the individual dislocations can be made; in some experiments edge dislocations can be distinguished from screw dislocations. The crystals most successfully used in such studies have been the alkali halides particularly LiF; with these, Gilman and Johnston have found it possible to distinguish pits associated with grown-in dislocations from those which have originated during plastic deformation and have subsequently moved to a new location (Review 3; Johnston, et al., 1959). With this technique many studies of dislocations have been made (generation, motion, velocity, dynamic resistance, shape of dislocation loops, etc.). While most of the LiF work has been done at low dislocation densities, recent work shows that in crystals with high dislocation densities it is possible to age these so as to distinguish between the aged ones and a few new ones introduced subsequent to the aging, thereby making it possible to study individual mobile dislocations even in a strain-hardened crystal (Gilman, et al., 1959).

When a dislocation is parallel to and near the surface of a crystal it is sometimes possible to reveal a considerable length of it by etching; special heat treatment to induce precipitation along the dislocation may be important in such work (Review 1). Although etch pit methods have been particularly effective with alkali halide and other inorganic crystals, it is clear that various refinements of the methods that have been worked out for these substances should be carried over to other materials—and this is being done to some extent.

Etching by thermal action at elevated temperatures, sometimes combined with oxidation or other chemical reaction and with evaporation, should be mentioned (sec. 2.10). The use of elevated temperatures naturally limits the applications of these methods. Other methods of etching may have advantages in some cases, e.g., ion bombardment (Holland, 1956) and electrolytic etching under controlled conditions (Review 1; Edeleanu, 1957).

The birefringence caused by the stress field surrounding individual dislocations, and surrounding arrays of dislocations in a slip band, seen by the use of polarized light, has been used in studies of dislocations in crystals transparent to visible light,

e.g., silver chloride, and crystals transparent in the infrared, e.g., silicon (Review 2). To interpret these polarized light patterns best, the light must be parallel to the dislocations in a thin sheet, or at least to a plane of dislocations. This requirement constitutes a severe limitation of the method in general studies of plastic flow; the resolution of individual dislocations in complex arrays is usually beyond the capabilities of the method.

In many crystals it is possible to make dislocations visible over their entire length by decorating them, i.e., by causing small particles to be deposited along their length (sec. 2.10). The deposition procedures require some annealing and thus limit the types of problems that can be approached in this manner, as does the fact that the crystals must be transparent to visible or infrared light; nevertheless, the amount of information that has been acquired through the method is extensive (Review 3).

2.1.2. X-Ray Methods

The X-ray method of estimating dislocation densities by using the distortion of Laue and Bragg spots, the widening of diffraction lines, measurement of integrated intensities, and analysis of diffraction line contours do not reveal individual dislocations (Reviews 1, 2; Hirsch, 1956; Guinier, 1952; Barrett, 1952).

When a fine-grained photographic plate is placed very close to a reflecting specimen, the recorded image of a Bragg reflection, when enlarged, can reveal much detail regarding the imperfections in the specimen (Review 2; Barrett, 1952; Barrett, 1945; Newkirk, 1959). It has become clear that individual dislocations appear in these images (Newkirk, 1959) as well as subgrains, deformation bands, and bands of segregation (Barrett, 1945). Some information on the direction of Burgers vectors of the dislocations can be obtained, and stereoscopic viewing of the photographs is possible. The depth of penetration of the rays into the surface is sometimes a limitation, for this is frequently quite small. The use of still-finer grained emulsions, closer distances between specimen and plate, and smaller X-ray focal spots should increase the resolution of the method beyond the one to five microns normally attained. The use of developers that avoid solution and redeposition of silver, and perhaps emulsions coated directly on the surface of the specimen might improve resolution. If the incident beam is reflected from a monochromator crystal of high perfection and then from the specimen in a double crystal spectrometer, the angular distortion around a dislocation can be investigated in a precise way (Bonse and Kappler, 1958). The method

has not yet been tried extensively in conjunction with techniques that decorate dislocations.

For crystals that are not too absorbing for the radiation used, the reflections can be recorded after transmission through a crystal slab. Lang (1959) has photographed the course of individual dislocations, clusters of them, and subgrain boundaries throughout the interior of crystals in an elegant technique by using a narrow sheet of rays striking a crystal and by having crystal and film shift together under the beam of X-rays. The photographs resolve dislocations 5 to 20 microns apart and should be capable of identifying the directions of Burgers vectors if different reflecting planes are used on a given sample.

A method related to that of Lang's employs the direct beam rather than the reflected beam. In a crystal of high perfection, the beam is transmitted with an anomalously low absorption coefficient [the Borrmann effect (Borrmann, 1941, 1958)], except where it encounters imperfect regions such as dislocations. Thus, dislocations cast a shadow in the beam, and their shadow can be recorded on a film (Lang, 1959), either with a stationary technique or one in which specimen and film are shifted together.

Hunter (1958) has pointed out that with a crystal of considerable perfection, intensity measurements of the transmitted beam as a whole, because of the Borrmann effect, are capable of detecting dislocation densities as low as tens or hundreds per square cm. Inhomogeneous strains from causes other than dislocations are also detected—including, presumably, clustered vacancies. Therefore, interpretation cannot always be unambiguous, and since only the total beam is measured, the method is not an image-forming one. The use of beams smaller than the 1.0×0.1 mm beam originally used by Hunter might aid in interpreting the distribution of local strains, but it would be necessary to combine the intensity measurements with photographic mapping of the defects present if such a point-by-point exploration of a crystal were to yield any reliable information on local strains.

2.1.3. Electron Microscope Methods

The methods mentioned in the preceding sections have their counterparts in electron microscope work. Replicas of etch surfaces are effective, and a little has been done by surface-reflection electron microscopy, but the greatest amount of information comes from electron beams transmitted through thin specimens (Reviews 2-3). The resolution that can be obtained with the best microscopes is so fine, when conditions are favorable, that a further increase

in resolution would be of relatively trivial importance in observing dislocations and stacking faults. Exploitation of the potential of present methods and equipment may require many years, since the preparation of thin samples and composite samples is not easy, and since a full understanding of the details seen in the micrographs often necessitates additional theoretical work. The preparation of thin samples by deposition on various substrates, by electrolytic polishing, by mechanical thinning, or by other methods may be further developed. Further use may be expected of the technique of producing moiré patterns from superimposed layers of crystalline material. It has been recently noted, for example, that silver deposited in certain thicknesses on mica substrates is smooth on an atomic scale and can be used as a substrate for other deposits (Pashley, 1959); much work has been done with various epitaxial deposits on crystalline substrates.

Opportunities for electron microscope work exist in the (1), study of interactions of complete and incomplete dislocations with each other and with other defects, with grain boundaries, and with precipitates; (2), the clustering of point defects and their collapse into dislocation loops and stacking faults, the distribution of dissolved atoms and the details of early-stage precipitation; (3), the photography of super-lattice domains; (4), segregation at boundaries; and (5), conditions involving the nucleation of cracks (Review 2; Hirsch, et al., 1958; Kuhlmann-Wilsdorf, et al., 1958; Silcox, et al., 1958).

To lessen the limitations of the method that are inherent in the thinness of the samples required, it is possible to gain somewhat in penetrating power by going to higher voltages. Any method of specimen preparation that retains the dislocation distribution characteristic of a thick crystal through the thinning and photographing process deserves attention, for thin-crystal dislocation behavior is not always identical with that in thick crystals. Research is in progress on the straining of specimens during observation by the application of thermal stresses that are usually of unknown magnitude and direction, and also by the external application of stresses. Advances can be expected in which observations are made at controlled nonambient temperatures; with controlled distributions of second-phase particles, subboundaries, or grain boundaries. Detailed observations may provide important measures of the energies of fault-planes, grain boundaries, subboundaries, and superlattice domain boundaries.

There is a gap between high-resolution electron micrographs on the one hand and the micron-range resolution of the X-ray imag-

ing methods on the other; while this gap is closed to some extent by the etch pit methods, it would be advantageous to have additional methods that would help fill the existing void.

2.1.4. Point Emission Microscope Methods

The images formed on a fluorescent screen by electrons emitted from a small metal point when subjected to an electric field in a high vacuum show at very high magnification the distribution of emission over the surface of the emitter (Review 1). The images, with resolution reaching the neighborhood of 20Å, provide information on crystal structure and orientation, work function and its anisotropy, surface contamination with monatomic and polyatomic layers that are physically or chemically adsorbed, rates of surface diffusion, subgrains, twins, and a limited amount of information on dislocations that reach the surface or other irregularities of the surface. The points may be thinned-down wires or may be whiskers, and they may be subjected to severe stresses during observation by means of the electric field of the microscope (Review 1; Silcox, et al., 1958).

The field ion microscope is a newer development based on a similar principle but permitting higher resolution and employing higher field strengths (Review 2; Gomer (in press); Good, and Müller, et al., 1956). Because of the high stresses from the electrostatic field the field ion microscope appears to be limited to the study of about ten high melting point metals.

Among the details that have been seen in this microscope by Müller and associates are the edges of layers of atoms, the emergence of slip bands, dislocation sources, changes accompanying fatigue of the point caused by pulsing the applied voltage, and apparently even individual lattice vacancies at the surface or brought to the surface by stripping atomic layers from the point. For further comment see Part IX.

2.2. Microradiography and Autoradiography

Techniques for microradiography have long been established and the field is not in a state of flux at present. The choice of suitable wave lengths to bring out needed contrasts in microradiographs is important. Microscopic variations in composition of a sample can be studied, at least qualitatively and in favorable cases also quantitatively, with microradiographs (Maddigan, 1946). Stereographic pairs of microradiographs have been used on metals to present the three-dimensional shapes of grains; to bring out grain boundaries, a metal was chosen that could be made to flow along

the boundaries, and that provided sufficient contrast with the material of the grains themselves (Williams and Smith, 1952). Some evidence of individual grains apart from concentration effects is also seen in X-ray microradiographs simply because of the removal of varying amounts of the transmitted ray by diffraction. Autoradiography may be employed in studying micro-distributions of elements that are radioactive, as for example, penetration of an element along grain boundaries. The method is of greater value for qualitative work than for quantitative work. Development work might lead to somewhat higher resolution than heretofore attained.

2.3. Small-Angle X-Ray Scattering

The distribution of intensity of scattered X-rays in the range of angles within a few degrees of the direct beam has attracted much interest because of the possibility of deducing, from the measured intensity distributions, considerable information regarding heterogeneities on a submicroscopic scale that exist in samples (Guinier and Fournet, 1955; Guinier, 1959). The inhomogeneities that are detected are those in the electron density of the scattering material; the interpretation is most rigorous when these consist of separate, well separated particles or precipitates, but continuous variations in electron density in a solid solution can also be studied. The early stages of age hardening of alloys produce important effects, since solute atoms become grouped in ways that differ markedly from the distribution of a random solid solution; much valuable information on the size and shape of clusters of atoms has been obtained from the data. The method has also been used to show the change of shape of small precipitated particles after deformation and to study catalysts, colloids, macromolecules and fibrous materials.

Successive reflection of rays within a sample (double Bragg scattering) is a major troublemaker in the small-angle region (Webb and Beeman, 1959) for it sometimes contributes more intensity than the inhomogeneities that are being studied. This is the problem, in general, that makes for difficulty in studies by this method of defects introduced into a metal by cold work. Another cause of small-angle scattering that can be guarded against is the presence of irregularities on the surface of the specimen, such as could be caused by oxidation. Uncertainties in interpretation enter when there is a distribution of particle sizes or shapes, and when there is inter-particle interference.

Low-angle scattering together with other X-ray methods of detecting defects (line broadening, diffuse scattering, integrated

intensities) have recently been reviewed in connection with problems of radiation damage (Simmons and Balluffi, 1959). The field is complex; various proposed models of the damage are usually compared with one or more sets of data, and it has not, in general, been possible uniquely and rigorously to determine the details of the damaged state from the observations.

2.4. Diffuse Scattering

The maximum information from X-ray scattering regarding submicroscopic inhomogeneities can be obtained by a complete analysis of the diffuse scattering throughout a representative portion of a unit cell in reciprocal space. The principal features of this scattering have given much information on the Guinier-Preston zones in early stages of age hardening and on the presence of stacking faults (Guinier, 1956; Averbach, 1956; James, 1948; Guinier, 1959). More quantitative intensity measurements, combined with careful corrections for Compton scattering and thermal diffuse scattering, are capable of providing additional information on clustering of atoms in solid solutions, on short-range order, on solid solution lattice distortions arising from atoms that differ in size, and on the distortions around isolated and clustered point defects (Guinier, 1959). (Strains around dislocations are probably not effectively attacked by this method.) Notwithstanding the success of electron microscopy in these fields, and the experimental difficulties of the diffuse scattering work and the uncertainties of corrections and interpretations, it can continue to contribute information not otherwise obtainable (Guinier, 1959).

The dispersion curves for the elastic waves in crystals can be determined from the thermal diffuse scattering, and interpreting these curves in terms of the Born theory of lattice vibrations yields values of the force constants between atoms; in a recent determination the nine force constants for first- second- and third-neighbor interactions in aluminum were determined, and from these were computed eigenfrequencies, the frequency spectrum with its singularities, and specific heats (Walker, 1956). Anharmonic terms are ignored in this analysis and in the Born theory so the results are approximate ones; the problems of including these terms have not been solved. Lattice vibration problems can be more directly approached by measurements of the wavelength distribution of inelastically scattered neutrons than by X-rays, so the X-ray approach is not very active at present in this field.

2.5. Precision Structure Determinations

Relatively few measurements of cell dimensions or of structure parameters have been made at very low temperatures where maximum precision should be possible, although techniques are well developed (Peiser, 1955). The intensities of high-order reflections are increased and more reflections become available for measurement at liquid nitrogen and liquid helium temperatures. Corrections of the integrated intensities for background scattering become less and the Debye-Waller temperature factor becomes a much smaller and much less uncertain correction at these temperatures.

In many cases only the most precise structure determinations are capable of convincingly proving or disproving any effects on the structures of alloys that may result from the Fermi surface approaching or overlapping the boundaries of a Brillouin zone. (For a review of possible effects see Massalski, 1958). The X-ray determinations must also be carried out on samples prepared with great care with regard to purity, homogeneity, and chemical analysis. Certain changes of slope in published curves of lattice constant vs. composition in some alloys, ascribed to the interaction of the Fermi surface with Brillouin zone boundaries, have not been confirmed in recent precision determinations (Walker, et al., in press), although such effects do appear to be present in other alloys. Experiments on alloys of critical compositions carried out at different temperature should be useful in deciding some questions in this area.

Precision lattice dimension determinations over a range of temperatures are being used in connection with dilatometric measurements to determine the equilibrium concentrations of point defects in metals (Simmons and Balluffi, 1959). Constancy of temperature of the sample throughout the length being investigated is a severe requirement in this work and relatively few studies have been completed. The question of the interpretation of observed lattice expansion and contraction in terms of interstitials, vacancies, dislocations and mixtures of these has received theoretical study; conclusions can be drawn from such data, but only by relying on supporting evidence or theoretical calculations as to the nature of the defects present and estimates of the relaxation of the strains around each type of defect or assumptions as to the law of expansion vs. temperature in the absence of defects. It would be desirable, of course, if these various uncertainties could be further reduced or could be eliminated.

A special method of lattice constant determination should be mentioned because its potentialities have not been widely exploited, yet it is capable of high precision. This is the use of Kossel lines (Barrett, 1952; Kossel, et al., 1935, 1937) that are formed when a source of rays diverges from a point source located near the surface of a crystal. The early work with this method has been reviewed some years ago (Lonsdale, 1937; Geisler, et al., 1948); extensions of the method are being explored more recently in connection with the distortions that occur in an alloy during age hardening. Since patterns of Al-3.85 percent Cu single crystals yield divergent beam patterns in which some lines are broadened more than others during early stage precipitation, the method appears to be capable of supplying information on the anisotropic strains around precipitates (Imura, et al., 1959).

The interpretation of integrated intensities of reflections is always pushed to the limit (or beyond) in attempting to determine interesting details of the electron density distributions in crystals, for example the distribution of bonding electrons and the state of ionization of atoms. For attempts of this type and for general use in crystal structure determination, there is need for determining accurate atomic scattering factors. The calculation of atomic scattering factors with the precision that now is possible should be encouraged. It would be desirable, also, if the reliability of dispersion corrections, temperature factors and extinction corrections were increased by additional theoretical and experimental work. The dependence of extinction on dislocation density has not been explored in many substances, but it has been shown to be negligible for CuK-radiation on germanium when the dislocation density is less than 10^5 per sq. cm (Batterman, 1959).

2.6. X-Ray Spectroscopy

Emission and absorption spectroscopy provide certain information on electron energy states in solids—the occupied states of the valence band in the case of emission, and the unoccupied states in the case of absorption. Various changes in the details of emission and absorption spectra have been found that depend upon chemical combination or alloying of the elements in a sample. The problems of instrumentation and of interpretation are extensive and have been treated at length in a number of reviews (Parratt, 1959; Sandstrom, 1957; Shaw, 1956; Parratt, 1959).

Information on the characteristic features of the curve of density of states (the $N(E)$ curve), Brillouin zone filling and overlap, and energy gaps between bands have been deduced from

experimental data, but not without some discrepancies between different observers and some uncertainties in interpretation. A complication in interpretation arises from the existence of local discrete states that occur at an atom in a solid in which there is a vacancy among the inner electrons, for the spectra arising from these states are superimposed on the spectra from the energy bands of the normal, unperturbed solid. This complication has frequently been ignored in drawing general or preliminary conclusions, but will receive increasing attention in the future. The problems of separating the local discrete state spectra from the unperturbed band spectra are largely unsolved.

For this separation to succeed and for reliable information to be deduced regarding the local states themselves, which are of interest in solid state physics, many careful corrections must be applied to the experimental data (some of these are frequently ignored even though very large) and the data must be acquired using the best techniques; on the theoretical side, various simplifying assumptions have been made that have effects of unknown magnitude on the final conclusions and that require further study (Parratt, 1959). Corrections must be made for nonlinearity of the intensity scale, instrumental resolving power, self-absorption by the material emitting the X-rays, background under the emission band, satellite emission lines, and width and shape of the high-energy state. Care must be taken to avoid surface contamination (oxidation or other reactions, deposits of tungsten or carbon, etc.), and high surface temperatures that might induce phase changes or other alterations. When the sample investigated is not a pure element, several of the corrections must be worked out individually for each element present. If the $N(E)$ curve is desired for material unperturbed by single or multiple inner-electron vacancies, the corrections for excitation spectra and transition probabilities must be considered. Since each energy level or band contains a mixture of wave-function symmetries, the details of which are generally unknown, these last two corrections make the derived curve of $N(E)$ vs. energy E less reliable than the curve without these two corrections, namely the curve $N_a(E)T(E)$ vs. E , i.e., the product of the transition probabilities $T(E)$ with the density $N_a(E)$, of those states (including excitation states) involved in the given transition.

Other X-ray methods for studying energy bands in solids have been explored: (a) the intensity of an emission line as a function of applied voltage when the voltage is varied through a range of values just sufficient to raise an inner electron to various unoccu-

pied states (Nilsson, 1958), and (b) the curve of intensity vs. wave length for the continuous radiation near the short wave length limit (Stephenson, 1957), a curve which reveals the influence of the unoccupied levels into which the incoming electrons from the cathode fall. Neither of these methods avoid complications from multiple inner electron vacancies or from plasma oscillations (Parratt, 1959). A new development in the field is the high intensity source of X-rays discussed below.

2.7. High Intensity X-Ray Sources

X-ray spectroscopy in the future may see increasing use of a new source of continuous ("white") radiation that provides high intensity in the long-wave length region beyond 30 or 40Å. The source is the centripetal acceleration of electrons circulating in a synchrotron (Tombouliau and Hartman, 1956; Parratt, 1959). The radiation is concentrated in a rather narrow beam that can be intense from 50 to 300 or 400Å. This source is suitable for transmission through thin absorbers in absorption spectroscopy or for the excitation of fluorescence radiation for emission spectroscopy. It has also been suggested as a source to be directed through very fine pinholes at a specimen in a microbeam X-ray fluorescence analysis apparatus, competing with the electron probe method.

High intensity X-ray tubes utilizing rotating anodes have had a long history of development, as have also the tubes used for flash radiography. It would be unwise to conclude, however, that other design possibilities have been exhausted, particularly for tubes needed for some particular type of experiment. Perhaps the limitation on power input that results from volatilization of the target could be circumvented by designs that somehow permit volatilization. The ratio of characteristic to "white" radiation might be improved by employing the rays that leave the target on a line with the electron beam, directly backwards through the filament. Pulse excitation of a tube together with synchronized circuitry in the detecting counter could offer advantages in the signal-to-noise ratio for a given average power input to an X-ray tube.

2.8. Electron Probe X-Ray Microanalysis

At present the resolution possible in X-ray microanalysis using electron probes of the Castaing type (Castaing and Guinier, 1950) is limited principally by the penetration and scattering of the electrons in the target. Theoretically, it would be possible to make the volume in which the characteristic X-rays are excited infinitely

small by making the accelerating voltage arbitrarily close to the critical excitation potential. However, since the intensity of the continuous background remains finite for accelerating voltage equal to the critical excitation potential, there exists an optimum ratio of accelerating voltage to critical excitation potential for practical analysis.

Another possibility for improving the resolution of electron probe microanalysis would consist in making use of progressively longer wavelength emission radiation. For example, one could use the L-radiation of elements below atomic number 50 and the M radiation of heavier elements. Present technology of soft X-ray spectroscopy has a practical limit of about 10Å using organic crystals and windowless flow proportional counters, but by utilizing gratings and special detectors it may be possible eventually to extend this to 50Å or more. Assuming the development of appropriate techniques for analyzing X-radiation of 10 to 50Å, the resolution then becomes limited chiefly by the size of the electron focal spot. In practice electron focal spots of the order of 100Å in diameter have been achieved. However, the current obtained on such focal spots is very small, since the current decreases more rapidly than the area of the spot (proportional to $d^{3/2}$) if the electron source is a thermionic emitter. Therefore, from the standpoint of obtaining an adequate signal it is likely that 0.1 micron will represent the best resolution obtainable in practice.

2.9. References

- Averbach, B. L., 1956, *Am. Society of Metals*, 301-319.
Barrett, C. S., 1945, *Trans. Am. Inst. Min. Met. Engrs.* 161, 15.
Barrett, C. S., 1952, *Structure of Metals*, McGraw-Hill, New York.
Batterman, B. W., 1959, *J. Appl. Phys.*, 30, 508.
Bonse, V., and Kappler, E., 1958, *Zeit Naturforschung*, 13a, 348.
Borrmann, von G., 1941, *Physik. Zeit*, 42, 147.
Borrmann, von G., Hartwig, W., and Irmeler, H., 1958, *Z. Naturforschung*, 13a, 423.
Castaing, R., and Guinier, A., 1950, *Proceedings on a Conference on Electron Microscopy at Delft*, 1949.
Edeleanu, C., 1957, *Jour. Iron and Steel Inst.*, 185, 482.
Geisler, A. H., Hill, J. K., and Newkirk, J. B., 1948, *J. Appl. Phys.*, 19, 1041.
Gilman, J. J., and Johnston, W. G., 1959, private communication.
Gomer, R., *Field and Ion Emission*, Harvard Univ. Press, Cambridge, in press.
Good, R. H., and Müller, E. W., 1956, *Handbuch der Physik—Encyclopedia of Physics*, Vol. 21, 174.
Guinier, A., 1952, *Imperfections in Nearly Perfect Crystals*, Wiley and Sons, New York, 402-440.
———, and Fournet, G., 1955, *Small Angle Scattering of X-rays*, Wiley and Sons, New York.

- , 1956, *Trans. Am. Inst. Min. Met. Engrs.*, **306**, 673.
- , 1959, *Solid State Physics*, Vol. 9, 293–398, 294–400.
- Hirsch, P. B., 1956, *Progress in Metal Physics*, Vol. 6, 236.
- , Silcox, J., Smallman, R. E., and Westmacott, K. H., 1958, *Phil. Mag.*, **3**, 897.
- Holland, L., 1956, *Vacuum Deposition of Thin Films*, Chapman and Hall, Ltd., London, 439–443.
- Hunter, L. P., 1958, *Proc. Ser. B.*, **61**, 214.
- Imura, T., Nakayama, Y., and Weissman, S., paper presented at the Pittsburgh Diffraction Conference in 1959.
- James, R. W., 1948, *The Optical Principles of the Diffraction of X-rays*, G. Bell and Sons, London.
- Johnston, W. G., and Gilman, J. J., 1959, *J. Appl. Phys.*, **30**, 129.
- Kossel, W., Loack, V., and Voges, H., 1935, *Zeit. Physik*, **94**, 139; 1937, *Ergeb. exakt. Naturwiss.*, **16**, 296.
- Kuhlmann-Wilsdorf, D., Maddin, R., and Kimura, H., 1958, *Z. Metallkunde*, **49**, 584.
- Lang, A. R., 1959, *Act Cryst.*, **12**, 249, 1959, *J. Appl. Phys.*, **30**, 1748.
- Lonsdale, K., 1937, *Phil. Trans. Roy. Soc.*, **A240**, 219.
- Maddigan, S. E., 1946, *Industrial Radiography*, **4**, 22–25.
- Massalski, T. B., 1958, *Metallurgical Reviews*, **3**, 45.
- Newkirk, J. B., 1959, *Trans. Am. Inst. Min. Met. Engrs.*, **215**, 483–496.
- Nilsson, A., 1953, *Arkiv Fysik*, **6**, 513.
- Parratt, L. G., 1959, *Rev. Mod. Phys.*, Vol. 52, 616.
- , 1959, *Solid State Physics*, Vol. 6, 281–292.
- , 1959, *Rev. Sci. Inst.*, **30**, 297.
- Pashley, D. W., 1956, *Proc. Roy. Soc.*, **A210**, 355; 1959, *Phil. Mag.*, **4**, 316.
- Peiser, H. S., 1955, *X-Ray Diffraction by Polycrystalline Materials*, The Institute of Physics, London.
- Sandstrom, A. E., 1957, *Handbuch der Physik—Encyclopedia of Physics*, Vol. 30, 246–304.
- Shaw, C. H., 1956, *Am. Soc. for Met.*, 13–62.
- Silcox, J., and Hirsch, P. B., 1958, *Phil. Mag.*, **4**, 72.
- Simmons, R. O., and Balluffi, R. W., 1959, *J. Appl. Phys.*, **30**, 1249.
- Stephenson, S. T., 1957, *Handbuch der Physik—Encyclopedia of Physics*, Vol. 30, 368–370.
- Tomboulion, D. H., and Hartman, P. L., 1956, *Phys. Rev.*, **102**, 1423–1447.
- Walker, C. B., 1956, *Phys. Rev.*, **103**, 547.
- , and Marezio, M., "Lattice Parameters and Zone Overlap in Solid Solutions of Lead in Magnesium," *Acta. Met.*, in press.
- Webb, M. B., and Beeman, W. W., 1959, *Acta. Met.*, **7**, 203.
- Williams, W. M., and Smith, C. S., 1952, *Jour. Metals*, 755–765.

2.10. Reviews

1. *Methods of Experimental Physics*, L. Marton, Ed. (Academic Press, New York, 1959), Vol. 6, Part A. Articles by S. Amelinckx, pp. 321–357, and by J. W. Faust, Jr., pp. 147–176, and by H. J. Yearian and J. F. Radavich, pp. 176–186.
2. *Internal Stresses and Fatigue in Metals*, G. M. Rassweiler and W. L. Grube, Ed. (Elsevier Publishing Company, New York, 1959). Articles by C. S. Barrett, pp. 15–40, A. Guinier, pp. 83–97, P. B. Hirsch, pp. 139–162.

3. *Dislocations and Mechanical Properties of Crystals*, J. C. Fisher, et al., Ed. (Wiley and Sons, New York, 1957).

3. HIGH PRESSURE, HIGH TEMPERATURE

3.1. Static Pressures and Temperatures Obtainable Today

Within the past five years, laboratory high-pressure, high-temperature apparatus has been developed to the point where pressures of 200,000 atmospheres may simultaneously be obtained with temperatures of 10,000°C. Pressures of this magnitude may be maintained for many hours. Temperatures of 10,000°C, however, may be maintained for only a few seconds. More conservative temperatures, of the order of 2,500°C, may be maintained for many hours as long as means are provided for cooling critical components of the high-pressure device.

3.2. Pressure Media

Since progress in the field of high pressures at high temperature is directly dependent upon the state of the art (in regards to apparatus) a few remarks concerning apparatus development are apropos. First, let us consider the means used to heat a specimen and also the material used to transmit pressure to the specimen. The most satisfactory method for heating a specimen to high temperature is by means of some type of electrical resistance heating element located inside the pressure chamber. With such an arrangement, it becomes necessary to insulate the walls of the confining vessel from the high temperature. Sir Charles Parsons (1888), in his early diamond synthesis studies, utilized a solid refractory medium for this purpose. In his apparatus, this solid material surrounded the electrical resistance furnace and the sample contained therein and also transmitted pressure to the heating element and sample. In the past there has been considerable hesitancy to utilize a solid material to transmit pressure. It is natural to assume that there is considerable frictional "hold-up" in a solid material and therefore poor transmission. At pressures greater than 35,000 atmospheres (at room temperature) *all* liquids that were liquid at 1 atmosphere and room temperature have become solid. (A 50-50 by volume mixture of n-pentane and isopentane remains liquid to 35,000 atmospheres. All other liquids freeze at lower pressures.) Hydrogen and helium would undoubtedly be liquids in the 100,000 atmosphere range of pressure but, at the present time, have not been used as pressure media because of difficulties encountered in confining these elements at very high

pressures. Media used to transmit pressures at room temperature, above 35,000 atmospheres, must of necessity, therefore, be solids. Consequently, it has become important to investigate the question as to how well various solid materials transmit pressure within a confined volume. Some work in this connection has been reported (Bridgman, 1952; Hyde, 1957; Cho, 1958), and experience has shown that frictional holdup is much lower than originally anticipated with many types of solid materials. Fine-grained compacts such as compressed powders of MoS_2 , BN, cornstarch, asbestos, mica or naturally-occurring pyrophyllite and talc have been found to be good media for transmitting pressure. The use of solids to transmit pressure has simplified high-pressure, high-temperature apparatus design and opened up new design possibilities. Frictional hold-up in some designs utilizing solid media have been as low as 3 percent (Boyd, 1958). In any event, frictional hold-up is reproducible and determinable.

At pressures to 35,000 atmospheres and temperatures to $1,500^\circ\text{C}$, liquids or gases have been used to transmit pressure to a sample contained in a tube furnace (Butuzov, et al., 1953; Birch, et al., 1957). With such systems heat loss is considerably greater than it is in systems that utilize a solid pressure medium and it is necessary to use baffles and reflectors to prevent the walls of the confining vessels from becoming excessively hot.

3.3. Conventional Apparatus

The usual high-pressure apparatus consists of a cylindrical vessel in which one end is fitted with a closure and the other end is fitted with a movable piston. A sample, confined by the walls of the containing vessel, the closure, and the piston, is compressed by the advancing piston. Electrical leads for making measurements and for providing heating power to a furnace confined within the device can be brought in through conical seals. Such devices when properly constructed can reach pressures in the neighborhood of 50,000 atmospheres. Failure of these devices centers around the piston. Cemented tungsten carbides are used for pistons and if materials with compressive strengths higher than those of cemented tungsten carbide were available, correspondingly higher pressures could be obtained. Troubles connected with piston and cylinder devices are those of closure (gasketing and sealing) and the friction and binding of the piston as it moves in and out of the cylindrical chamber.

Certain "tricks" must be used to reach pressures beyond the 50,000 atmospheres that are obtainable in simple piston and cyl-

inder devices. The strength of a right circular cylinder in compression depends (at least to pressures of several hundred thousand atmospheres) on the difference between the axial pressure and the pressure perpendicular to the axis. Thus, a cylindrical carbonyl piston surrounded by a fluid at one atmosphere when subjected to an axial load of about 50,000 atmospheres will fail. However, if the same piston is surrounded by a fluid at one atmosphere when subjected to an axial load of about 50,000 atmospheres it will be able to support an axial load (force per unit cross sectional area) of 100,000 atmospheres. As a matter of fact, some experiments performed by Bridgman (1952) indicated some strengthening of material subjected to such hydrostatic pressure, and this makes it possible for the axial load to be increased even beyond that indicated.

The above consideration shows that it would be possible by "multi-staging" (a process whereby one pressure apparatus is contained inside another pressure apparatus, and so on) to reach extremely high pressure. In practice, a three-stage apparatus has never been built because of mechanical complications. Bridgman, however, has successfully operated a two-stage apparatus at room temperature at pressures of 100,000 atmospheres and has made many useful measurements in such a device. The mechanical complications in this device are rather formidable, however, and it has not been possible to insert electrical leads into the inner chamber. There are also disconcerting factors such as the fact that piston-wall friction in the inner chamber may make up 40 percent of the load exerted on the piston.

3.4. Massive Support Principle

Some types of apparatus, other than two-stage design, capable of reaching 100,000 atmospheres are based on the idea of obtaining a multi-stage effect in a single stage. They are based on the so called "massive support" principle of Bridgman's. The massive support principle is the principle operating when the small faces of two broad truncated cones (called anvils) are forced together along their line of centers. The force per unit area load withstood by the truncated portion of these cones before failure is much greater than the force per unit area load that could be sustained by a right circular cylinder of the same cross section area. This is possible because the anvil faces have mechanical ties fanning out behind the surface of the truncated cone into the broad base behind the cone. This is the principle of massive support. Thin wafers of material may be subjected to pressures between the faces of such

truncated cone-type anvils and may be prevented from extruding laterally by a ring-type gasket usually constructed of a fine-grained stony material such as pyrophyllite. Besides preventing lateral extrusion of the waferlike sample, the gasket material compresses with the advancing anvils thus permitting the sample to be compressed.

3.5. Pseudo-Multistaging

In addition to obtaining higher pressures by virtue of "massive support" it is possible to obtain a pseudo-multistaging effect by compressing solid materials between the sloping shoulders of the anvils. In such designs the space between the faces of the anvils is subjected to the highest pressure while the materials located between the sloping shoulders of the anvils provide a pressure gradient such that pressure decreases in going away from the anvil faces. This additional support given to truncated cone-type anvils makes such anvils useful at still higher pressures. A design embodying this principle has been described by Fitch, Slykhous and Drickamer (1957). This particular device is suitable for optical studies at room temperatures at pressures to 200,000 atmospheres.

3.6. Belt Apparatus

A high-pressure, high-temperature device known as the Belt apparatus that was devised by Hall in 1953, is capable of reaching pressures in the neighborhood of 200,000 atmospheres simultaneously with temperatures above 3,000°C (Hall, 1960). Another apparatus, the Tetrahedral Anvil apparatus, developed by Hall (1958a, 1958b; Hall and Kistler, 1958) at Brigham Young University, has been successfully used at pressures of 130,000 atmospheres and temperatures of 3,000°C. This device utilizes the principles of massive support and motion via a compressible gasket introduced by Bridgman.

3.7. Possibilities of Higher Pressures for the Future

The number of runs that can be made in a high-pressure device before failure is set by the pressure of operation. Several dozen runs may be possible before apparatus failure at 100,000 atmospheres, about six runs may be possible at 150,000 atmospheres and an average of only one run before failure at 200,000 atmospheres. The upper limit of pressure obtainable in a device is that pressure at which failure will occur on the average on each run. The upper limit of pressure of any given device is set by the materials of construction. The materials with the highest com-

pressive strengths available today (in large sizes) are the cemented tungsten carbides. When materials of higher compressive strengths become available the pressure obtainable in any device will increase in direct proportion to the increased compressive strength. Diamonds have the highest compressive strength of any known material. Although the cost and availability of massive, single crystals of diamond prohibits their use in most high-pressure apparatus, fine diamond powder is readily available and the cost is low enough that it could be used if it were possible to make a cemented diamond composite analogous to the cemented tungsten carbides. There has been a small amount of effort to make such a material. Certainly, additional effort in this direction would be worthwhile.

The principles of massive support and of motion via compressible gaskets have been the prime factors responsible for the attainment of the pressures available today. Maximum use of these principles seems to have been achieved in the Tetrahedral Anvil device. The principle of pseudo-multistaging as used in Drickamer's device (Fitch, et al., 1957) has, however, not been fully exploited and it is believed that experiments with this principle in connection with the Tetrahedral Anvil device would eventually allow pressures as high as 300,000 to 400,000 atmospheres to be obtained. If an engineering material such as a cemented diamond powder with a compressive strength of 1,250,000 pounds per square inch could be made it would seem reasonable to expect that pressures of 500,000 to 600,000 atmospheres could be obtained in a pseudo-multistaged Tetrahedral Anvil apparatus.

All of the above methods for obtaining static high pressures are based on mechanical principles; i.e., the movement of pistons or anvils. It is possible to generate static high pressures by means of (a) phase changes, (b) chemical reactions, (c) thermal expansions, or (d) the pull of gravity on a long column of liquid. These methods have not had widespread use because the theoretical maximum pressures obtained are not very high in a, b and c and because the mechanical problems in d are difficult to resolve. For example, the maximum theoretically obtainable pressure generated by freezing water is about 2,000 atmospheres, and that obtained by freezing bismuth about 18,000 atmospheres. However, the maximum theoretically obtainable pressure generated by the freezing of germanium is at least 180,000 atmospheres and the use of this material in this manner would be worth further investigation.

3.8. Effects to be Expected at High Pressure

In many respects the effects of pressure are diametrically opposite to the effects of temperature. For example, when a solid material is heated it may pass through certain phase changes, it will then melt, and then become a gas. At still higher temperature a molecular gas will dissociate into atoms which will at still higher temperatures become ions. This entire process is one in which the system "opens up" and the entropy is increased. On the other hand if pressure is applied to a gaseous system, atoms may be forced to collapse to molecules, additional pressure may liquefy the gas which may later become a solid and which then may go through phase changes until a close-packed configuration is obtained. This process is one in which the system is continually "closed up" and the entropy is decreased. As an example, the entropy of potassium metal decreases by about 30 percent on compression to only 12,000 atmospheres. When systems in closest-packed metallic arrangements are compressed, it is possible to obtain a discontinuous decrease in the volume without achieving a new atomic arrangement. This occurs because the pressure is affecting the electronic structure. Important examples of this are to be found in the elements cesium and cerium. The transition in cerium has been studied as a function of pressure and temperature. At room temperature the lattice constant a of the face-centered cubic cerium lattice is 5.14Å. At a pressure of about 7,000 atmospheres, a decreases rather abruptly to 4.84Å. This corresponds to a decrease in volume of about 17 percent. As the temperature decreases below room temperature the pressure required to cause this transition decreases until at 109°K the transition occurs spontaneously at one atmosphere with an accompanying volume decrease of 10 percent. This transition is thought to be caused by a shift of a 4*f* electron to the 5*d* level. This type of behavior—rare at pressures below 100,000 atmospheres—is expected to become common at pressure above 200,000 atmospheres and may become the most exciting aspect of high pressure work in the future. Since these extremely high pressures are capable of influencing electronic configurations, and since chemistry depends upon these configurations, it is possible that new and hitherto unsuspected types of chemical bonding might be obtained under very high pressure conditions. The hope that some of these new arrangements might show a degree of metastability when pressure is released is encouraged by the fact that the cerium transition is stable at one atmosphere at a sufficiently low temperature.

At a sufficiently high pressure all nonmetals could be made to pass through a metallic phase. This would happen when the kinetic energy of the electrons is raised sufficiently to allow them to escape from the electric fields of their nuclei. Calculations in this connection show that ammonia should become metallic at a pressure of about 250,000 atmospheres and that hydrogen should become metallic at pressures of about 450,000 atmospheres. At still higher pressures inner electrons as well as valence electrons would become detached from their nuclei, the individual identity of atoms would be destroyed and the material would become a homogeneous mass of nuclei and electrons. At the density of material existing in the centers of the white dwarf stars (10^{16} atms) these systems would be completely degenerate. Only the lowest accessible energy states would be occupied and in these circumstances the temperature would be effectively zero. At extremely high pressures the uncertainty principle becomes important because as matter is compressed into a smaller volume the position of individual particles is known with greater certainty. This must be offset by an increased uncertainty in the momenta of the particles and this can be accomplished only by an increase in their average kinetic energy.

3.9. Melting Phenomena

A pressure of 100,000 atmospheres can change the melting point of some substance by as much as $1,000^{\circ}\text{C}$. Pressure is probably the most important tool available for the study of melting phenomena and for the elucidation of the nature of the phenomena. There should be increased emphasis placed on studies of melting at extreme pressure with the aim of improving theories of melting.

3.10. Synthesis of New Materials

Just as an increase in temperature generally increases the rate of a chemical reaction, an increase in pressure generally retards the rate of a reaction because the diffusion of atoms is suppressed. In order for a chemical reaction to proceed at a measurable rate at pressures of 100,000 atmospheres, then, it is almost always necessary that the system be heated. In most instances temperatures above $1,000^{\circ}\text{C}$ are necessary to counteract the effect of high pressure in retarding reaction rates. This is the reason that it is so very important to have high pressure and high temperature at the same time.

The following new materials (stable or metastable under room conditions) have all been produced under high-pressure, high-tem-

perature conditions: black phosphorus, black solid carbon disulfide, dense silica (coesite), diamond, cubic boron nitride, all five naturally occurring garnets and several new garnets not found in nature, dense boron phosphate, dense boron arsenate, and many others. Research directed at obtaining new materials by utilization of high pressure at high temperature is certainly a worthwhile program for the future.

3.11. Geology

A pressure of 200,000 atmospheres corresponds to that existing at a depth in the earth of about 400 miles. Temperatures existing at that depth are thought to be lower than those obtainable in the laboratory. We now have a tool of obvious geological importance and certainly much will be learned in the near future concerning the nature of the interior of the earth by laboratory high-pressure, high temperature studies. Additional emphasis placed on such work would be very worthwhile.

3.12. Metallurgical Applications

High pressure can greatly influence solubilities, phases, voids and crystallite sizes in metal systems. It is possible to obtain metallic phases by application of high pressure that can be obtained in no other way. The use of pressure as a variable, in addition to the variables of composition and temperature which have been extensively utilized in metallurgy, requires additional emphasis.

3.13. Conclusions

Temperatures of 1,000°C introduce energy into condensed systems that may be comparable to the energy associated with chemical bonds. Since ordinary flames readily produce temperatures of 1,000°C, temperatures sufficient to cause chemical changes to take place have been available for many thousands of years. The pressure required to introduce energy into condensed systems comparable to that introduced by temperatures of 1,000°C is of the order of 100,000 atmospheres. Although pressure is one of the oldest of the natural variables, such pressures have become available in the laboratory only within the past 15 years, and have been available simultaneously with high temperatures for only about five years. Consequently, research under such conditions is just getting started and it seems safe to predict that much will be learned in the near future from work in this field. It is evident that there will be contributions to most branches of the physical science (including geology and astronomy).

3.14. References

- Birch, F., Robertson, E. C., and Clark, S. P., Jr., 1957, *Ind. & Eng. Chem.*, **49**, 1965.
- Bridgman, P. W., 1952, *Am. Acad. Arts Sci. Proc.*, **31**, 165.
- Boyd, F. R., 1958, Geophysical Lab., Wash., D.C., private communication.
- Butuzov, Gonikberg, and Smirnov, 1958, *Doklady Akad. Nauk SSSR*, **89**, 651 (NSF-tr-76).
- Cho, Y., 1958, "Propagation of High Pressure in the Solid State," MA Thesis, Brigham Young University, Dept. of Physics.
- Fitch, R. A., Slykhous, T. E., Drickamer, H. G., 1957, *J. Opt. Soc. Am.*, **74**, 1015.
- Hall, H. T., 1958a, *Rev. Sci. Instr.*, **29**, 267.
- , 1958b, *Science*, **128**, 445.
- , 1960, *Rev. Sci. Instr.*, **31**, 125.
- , and Kistler, S. S., 1958, *Ann. Rev. Phys. Chem.*, **9**, 395.
- Hyde, G. R., 1957, "Friction at Very High Pressure," M.S. Thesis, Brigham Young University, Dept. of Physics.
- Parsons, Sir C. P., 1888, *Proc. Roy. Soc. (London)*, **44**, 320; 1919, *Trans. Roy. Soc. (London)*, **220A**, 67. See also Threlfall, Richard, 1909, *Engineering* **87**, 425.

4. HIGH TEMPERATURE

High temperature technology has made tremendous progress during the last ten years. In electric arc plasma-jets, temperatures can be produced and maintained for relatively long times, up to about 25,000°K. Shock tubes, exploding wires, and nuclear reactions are capable of creating still higher temperatures, for times in the range of milli- to microseconds. Techniques for producing high temperature per se are therefore available. For studying materials, however, the conditions present at high temperature are of the utmost importance, and the real difficulties in performing meaningful experiments at high temperature arise from the interaction between the material under study and its environment. Under certain conditions, these difficulties may be serious at relatively low temperature and new and improved techniques are needed for studying materials at 2,000°K. as well as at 20,000°K.

In the following, an attempt is made to review the basic limitations of the various methods for producing high temperature. For convenience, the discussion has been divided into sections, according to the type of heat source.

4.1. Electrical Resistance and Induction Heating

Electric heaters constitute the simplest and most practical type of heat source for the study of materials. The maximum tempera-

ture that could be obtained by this technique is limited by the temperature to which the heating element can be raised. The most refractory solid material known so far, namely a solid solution of tantalum carbide-hafnium carbide, has a melting point of about 4,215°K. Allowing for a certain margin between service temperature and melting point, it appears that 4,000°K is about the highest temperature that will be attained with electrical resistance heating elements. Graphite is the most widely used high temperature electrical heating element and under atmospheric conditions, its sublimation temperature is the limiting factor. Because the vapor pressure of graphite is about 1 atmosphere at 3,900°K, its use in electric furnaces is limited to about 3,600°K. There is little hope for increasing this practical maximum temperature by pressurizing the furnace, since the triple point of graphite is only about 4,000°K at a pressure of 100 atmospheres (Kohl, 1960).

Heating elements made of either carbides or graphite must be operated in vacuum or in an inert gas atmosphere. If an oxidizing atmosphere is required, the heating element, in addition to having a high melting point, must be inert to oxygen. By using thorium oxide, with a melting point of about 3,400°K (highest melting point of all oxides), it would be possible, at least in principle, to reach about 3,200°K. A furnace with thorium oxide (containing a certain amount of rare earth oxide) has been built by Lang and Geller (1951) and operated satisfactorily around 2,300°K. With further development, this temperature might be increased to about 3,200°K but there is not much hope for further improvement. Although resistance and induction heating processes are different in principle, the problems connected with the development of susceptors for producing very high temperature by induction are very similar to those encountered in the fabrication of resistance heating elements. In fact, the maximum temperatures that have been reached by induction were obtained with graphite susceptors, and the limitations mentioned in the previous section remain valid for induction heating.

Before discussing the other methods for producing high temperature, it should be pointed out that in spite of the fact that these methods lead to temperatures much higher than those obtained by either resistance or induction techniques, they do not lend themselves to achieving black body conditions. Measurements of the physical properties of materials at high temperatures require black body conditions and either resistance or induction heating remain the most practical methods for achieving these conditions. As a consequence, the presently available techniques for the study

of materials under constant temperature black body conditions are limited to about 3,600°K for materials stable under vacuum or inert gas conditions, and to about 2,300°K for materials requiring an oxidizing atmosphere. Further developments will not increase these temperatures beyond 4,000°K for vacuum and inert atmosphere, and 3,200°K for oxidizing atmospheres.

4.2. Solar Furnaces and Refocused Electric Arcs

By concentrating solar radiation with a parabolic reflector or with lenses, it is possible to obtain a high heat flux on a localized area, the size of the focal spot depending on the size of the concentrator. Obviously, the maximum temperature obtainable by this process cannot exceed the apparent temperature of the sun, namely about 6,000°K. It has been shown that under ideal conditions (Hiester, et al., 1957) a solar concentrator could eventually produce about 4,800°K. The maximum temperature actually reached in existing solar furnaces is probably less than 3,700°K. The area over which the solar radiation is concentrated corresponds to the solar image having a diameter $1/107.3$ times the focal length of the paraboloid mirror. It is therefore evident that the size and cost of solar concentrators become prohibitive when the size of the focal spot exceeds a few centimeters in diameter.

For a focal spot not exceeding a few centimeters, the solar furnace can be advantageously replaced by a refocused electric arc. In this scheme, a high intensity electric arc is placed at the focus of a parabolic reflector and the radiation is refocused by a reflector of identical focal length. Since high intensity electric arcs can generate temperatures higher than the apparent temperature of the sun, the refocused electric arc technique could extend the range of the solar concentrator to about 10,000°K. The limitations discussed above concerning the size of the high temperature focal spot of a solar furnace still exist in the refocused arc technique, since under the best experimental conditions the size of the refocused beam cannot appreciably exceed that of the arc itself.

Neither the solar concentrator, nor the refocused electric arc constitute promising tools for the study of materials at high temperature under black body conditions. The only way to obtain black body conditions with radiant energy is to trap the energy into a cavity made of a solid material. The limitations imposed by solid materials and discussed in the previous section, do not depend on the way these materials are heated to their useful maximum temperature, and radiation heating of a black body cavity does

not provide any advantage over more conventional and more practical methods such as resistance or induction heating.

For experiments performed under non-black body conditions, both solar concentrators and refocused electric arcs offer definite advantages. Perhaps the most important physical factor characterizing solar or refocused electric arc furnaces is the high heat flux available. Another important advantage is the fact that radiation heating is free from any kind of contamination. The material under study may be placed in a glass container, and surrounded by high vacuum or any gaseous atmosphere, including oxygen. Concentrated radiation heating is also convenient for melting refractory substances without crucibles, and this technique may be extended to zone melting and to growing refractory single crystals.

4.3. Chemical Flames

By burning hydrocarbons with oxygen, temperatures in the range of 2,000 to 2,500°K can be easily obtained. With less conventional chemicals used as propellants in rocket engines, 3,000 to 4,500°K can be obtained in the flame. According to Kirshenbaum and Grosse (1956) the combustion of carbon subnitride (C_4N_2) could produce temperatures in the 5,000 to 6,000°K range. It is difficult to imagine that chemical reactions could produce much higher temperatures, since dissociation of the products of combustion cannot be avoided.

The use of chemical flames for high temperature material research is complicated by the interaction between the flame and the material. This interaction leads to difficulties in the study of the intrinsic properties of materials. In some cases, however, this interaction may be the main purpose of the experiment. The behavior of solid materials in rocket chambers, rocket nozzles, etc., is probably best understood in experiments in which the actual environment is duplicated, and chemical flames are therefore indicated.

4.4. Electric Arcs and Plasma Jets

The electric arc was probably the first really high temperature source used in chemistry. The carbon arc of Moissan (1905) is still widely used in essentially its original form in many industrial processes. The modern technology leading to ultra-high temperature electric arcs originated with the pioneer work of Gerdien and Lotz (1922, 1935) who demonstrated that temperatures as high as 50,000°K could be achieved in water-pressurized electric arcs. Since then, many different types of pressurized arcs have been studied

and plasma-jets in which the pressurizing fluid is a gas are now commercially available. Air, oxygen, nitrogen, hydrogen, argon and helium can be used depending on the type of application. Jets with temperatures in the range of 5,000 to 20,000°K can be maintained for long periods of time, and none of the previously described techniques can achieve comparable performance.

Although the plasma-jet constitutes a unique and most useful tool in materials research, it has some important shortcomings. In a plasma jet, the material under study is surrounded by a hot plasma moving at high speed. When placed in the jet, the material under study interferes with the flow and it seems highly improbable that a uniform temperature distribution could be achieved under these conditions. The chemical interaction, present in chemical flames, however, can be eliminated in plasma-jets by the use of inert gases such as helium or argon. Refractory metals and compounds which are highly reactive at high temperature can therefore be heated and melted without difficulty. Plasma-jets are already available for spraying tungsten and other refractory metals, and compounds such as carbides, borides, silicides, etc. The Verneuille technique of growing single crystals of oxides in an oxygen-hydrogen flame, has been extended to several refractory compounds other than oxides, by replacing the chemically active flame by an inert gas plasma-jet. Before such crystals were available, most of the physical properties of refractory compounds had not been determined accurately, because the measurements were performed on specimens of less than 100 percent density and very often containing a second phase introduced by the use of a binder in the material preparation. The study of single crystals of refractory compounds should prove most rewarding and will certainly lead to a better understanding of the nature of the atomic bonds in carbides, nitrides, borides and the like.

4.5 Electron Beams

The stopping of electrons (or more generally of any particle having high kinetic energy) by a solid or a liquid "target" can generate large amounts of heat within limited volumes and consequently create high temperature. In recent years, this principle has been applied to electrons in the kilovolt range. The conditions under which these electrons are generated and accelerated require a vacuum of the order of 10^{-5} mm of mercury, and the material to be heated is therefore subjected to these environmental conditions. This may be an advantage for some high temperature problems, and the melting of refractory metals is an obvious applica-

tion of the method. Because of the high vacuum involved, purification of the molten metal can be achieved, if the thermodynamic conditions are favorable for the volatilization of the unwanted impurities.

Because of the high vacuum limitations, the electron beam heating technique in its present form, will probably be limited to melting (and refining) refractory metals and perhaps some compounds. It is conceivable, however, that the same principle could be extended to high energy electrons (in the Mev range) and to heavy particles. In this case, the particles generated and accelerated in high volume, could pass through a thin (and probably cooled) window and be used for heating purposes outside the vacuum chamber. The cost involved in such a process would exclude any foreseeable practical application, but it might help in solving some fundamental research problems.

4.6. High Temperature and New Materials

Since new materials—in the sense of solids having a different crystal structure than that generally obtained by usual processes—have been synthesized under high pressure, high temperature has been often mentioned as another potential source for obtaining unusual structures in solids. In fact, metallurgists have been doing just that for centuries, and nonequilibrium structures obtained by cooling solid alloys rapidly from high temperature are the rule rather than the exception. Very little attention, however, has been given to the possibility of extending the concept of rapid cooling to the liquid or the vapor phase. It is indeed conceivable, that some of the atomic bonds existing in both liquid and vapor states could be retained in the solid state, if the rate of cooling were high enough to disturb the normal process of phase transition. Buckel (1954) has shown that amorphous alloys could be obtained by vapor deposition on a target cooled by liquid helium in high vacuum. Because of the large mean free path of the atoms in the "vapor" phase present under these conditions, the cooling rate is a rather vague concept. In any case, amorphous alloys were obtained by achieving a rapid transition from high to low temperature.

Rapid cooling from the melt may also lead to unusual metastable phases in alloys. Duwez, et al. (1960a) developed a technique for cooling molten alloys into thin foils (about 10 microns) and obtained a continuous series of solid solutions in silver-copper alloys, which are eutectic alloys under normal equilibrium conditions. Solid solution was also achieved in germanium-gallium anti-

monide eutectic alloys (Duwez, et al., 1960b). Using the same technique of rapid cooling, Duwez, et al. (1960c) and Klement (to be published) synthesized a new compound around the composition Ag_3Ge in the otherwise eutectic silver-germanium alloys. Finally, Klement, et al. (to be published) obtained an amorphous structure in a rapidly cooled gold-silicon alloy containing about 25 atomic percent silicon.

The preceding remarks may serve to emphasize the fact that extremely high temperatures may not be needed for the synthesis of new materials. A more important problem may be in the development of techniques capable of achieving extremely rapid rates of cooling from either the liquid or the vapor phases. Similar "quenching techniques" might be conceived for retaining structures that are known to exist under high pressure.

4.7. REFERENCES

- Agte, C., and Alterthum, H., 1930, *Z. Techn. Physik*, **11**, 182.
Buckel, W., 1954, *Z. Physik*, **138**, 136.
Duwez, Pol, Willens, R. H., and Klement, W., Jr., 1960a, *J. Appl. Phys.*, **31**, 1136.
———, ———, and ———, 1960a, *J. Appl. Phys.*, **31**, 1137.
———, ———, and ———, 1960c, *J. Appl. Phys.*, **31**, 1500.
Gerdien, H., and Lotz, A., 1922, *Wiss. Veroff, Siemens-Konz*, **2**, 489.
———, and ———, 1935, *Wiss. Veroff, Siemens-Konz*, **14**, 25.
Hiester, N. K., Tietz, T. E., Loh, E., and Duwez, Pol, 1957, *Jet Propulsion*, **27**, 507.
Kirshenbaum, A. D., and Grosse, A. V., 1956, *J. Am. Chem. Soc.*, **78**, 2020.
Kohl, W. H., 1960, *Materials and Techniques of Electron Tubes*, Reinhold Publishing Co., New York.
Lang, S. M., and Geller, R. F., 1951, *J. Am. Ceram. Soc.*, **34**, 193.
Moissan, H., 1905, *Compt. Rend. Acad. Sciences*, **115**, 1031.

5. HIGH INTENSITY STRESS WAVES

5.1. Introduction

Observations of the response of solid materials to a sudden application of stress have received considerable attention in recent years. This is due in part to the increasing realization that the mechanical properties of certain important engineering materials are extremely time-dependent under rapid loading conditions and in part to the development of impulsive loading techniques for studying a number of important phenomena at stress levels that exceed, by a considerable margin, those obtainable under quasi-static conditions.

Studies of this type may be roughly divided into one of five categories.

1. Mechanical properties where the loading times are of the order of milliseconds or greater. Delayed yielding in carbon steel and irreversible and reversible flow of elastomers are examples of materials with important time-dependent mechanical properties in this range of times.

2. Mechanical properties in materials where the significant loading times are of the order of microseconds or less. Most of the crystalline solids fall into this category.

3. Thermodynamic properties under hydrostatic pressure. This category includes the compression-volume data for solids and the pressure-induced phase transitions with their accompanying structural changes.

4. Fracture of ductile materials at high stress levels in the absence of large strains.

5. Structural and property changes in materials subject to intense stress waves.

5.2. The Rod Configuration

Studies of the stress-strain behavior in the millisecond loading time range involves tensile or compression loading of wire-like or rod-like specimens under conditions where the complications due to stress waves can be avoided. Attempts to move into the microsecond range are frustrated by the initiation of stress waves, the interaction of the stress waves with the lateral surfaces of the specimen, and the interaction of the stress waves with each other. It is in this area that the problems of technique and instrumentation arise. The techniques for studies in the millisecond range have been solved, due to the efforts of the many investigators, notably Clark and Wood (1949), Davies (1948), Kolsky (1953), and Bell (1959), who have made important contributions, both analytical and experimental, to the stress wave picture of a rod-like specimen impulsively loaded by compression at one end. Unfortunately, the difficulty of ascertaining the stress-time situation in a specimen of this type, and the complexity of the result, limit experiments of this type in their usefulness for the study of properties of solid materials.

5.3. The Plate Configuration

A very much simpler stress-time situation exists when the impulsive loading conditions involve the application of high pressures for short times to one surface of a plate-like specimen. Here the edge effects are relatively unimportant and large compression

stresses may be induced directly into the solid. Large tensile stresses may be induced indirectly by reflecting a compression wave from the free surface (Hopkinson, 1905). Pressures in excess of 1,000 kilobars may be obtained in this way. Observations of the detailed pressure-time profile of a shock front will yield data on (1) the compressibility of solids, (2) the dynamic yield of solids and (3) pressure-induced phase transitions. Similarly, observations on the maximum tensile stress possible under these loading conditions will yield data on the fracture of ductile materials in the absence of large deformations.

5.4. The Source of High Pressures

The pressure within a detonation wave traveling through a condensed high explosion field may be as high as 350 kilobars. A shock wave may be transmitted into a solid when a detonation wave interacts with the solid interface. Relatively recent developments in the art of preparing explosive charges have made large homogeneous blocks of high explosive mixtures available to the experimenter. Combinations of fast and slow detonating explosives have been developed which allow one to initiate detonation at a point and develop a wide, essentially plane, detonation front that is flat to within a millimeter or two. This technique, and the others that have been developed at the Los Alamos Scientific Laboratory, involve a substantial amount of high explosive and therefore require a substantial facility for experimental work. An improvement in this situation is possible if one resorts to driving projectiles against plate surfaces. The present limiting factor is the limited velocity that a projective can acquire with existing gun or cannon techniques. Very high pressures may be obtained by accelerating a plate that serves as a projectile (or driver plate) with high explosives. However, here again, substantial explosive charges are necessary. The development of a cannon-like apparatus for accelerating driver plates to the 1,000 to 15,000 feet-per-second velocity range would definitely assist this work.

5.5. Measurements of the Shock Wave Parameters

From the point of view of materials research, the most important parameter of a shock wave is its pressure profile. What is needed is a pressure transducer which will operate over a wide pressure range (from substantially less than 100 kilobars to several hundred kilobars) with the associated measuring apparatus that will follow the property change in the transducer with a time

resolution of 10^{-8} seconds or so. Similarly, direct observations of the density profile of a shock wave with a resolution of better than one percent would be important.

The Rankin-Hugoniot relations, which are based on the conservation of mass, momentum and energy across the front of an advancing shock wave, make it possible for one to obtain data on the pressure and density behind the shock front indirectly by measuring the velocity of the shock wave and the particle velocity within the wave (Rice, et al., 1958). These relations are independent of the details of the shock front and only depend on the condition of the material ahead of and behind the shock front (assuming that the pressure profile is independent of time). Shock velocity data are obtained by measuring the arrival time of the shock front at two points displaced normal to the front. The particle velocity is obtained by measuring the free surface velocity (which is usually assumed to be twice the particle velocity). These quantities are in the kilometers-per-second range with the former somewhat faster than the latter. The particle velocity associated with a shock wave is more sensitive to the pressure within the wave than the wave velocity itself and is, therefore, the most important of the two.

Shock-wave velocities have been measured primarily by the pin-contact technique (Minshall, 1955). Here a number of holes of different depths are accurately machined into the free surface of a plate. Pins with their ends near the bottom of the holes serve to determine the shock-wave velocity by recording the arrival time at several points with suitable electronic equipment. Similarly a second set of pins, arranged at various free-run distances from the metal-free surface, provide data on the free-surface velocity. Similar data may be obtained with photographic techniques (Walsh, et al., 1955). Here free-surface motion drives a shock-wave into a gas (usually argon) that is contained in a narrow gap between the free surfaces and a lucite block. The shock wave in the gas produces brilliant luminosity which provides light for a photographic record of the time of arrival of the shock wave at that portion of the free surface. A second gap is arranged on the assembly at a second position displaced from the first in the direction of travel of the shock wave. A sweep camera views the assembly through slits and sweeps the image in a direction normal to the slits at a predetermined speed. If a third flash gap is positioned at a predetermined free-run distance from the free surface, the free-surface velocity may be determined.

The pin-contact technique has been used extensively by workers

in the field. When extremely well-developed it has a space resolution of 10^{-4} cm. The elastic-plastic two-wave structure of metals, as well as a two-wave structure due to a pressure-induced phase transition in iron, have been observed in this way. Unfortunately, due to the limited space resolution possible with this technique, it is very difficult to study the details of the wave structure. The flash-gap technique has been used with great success by the Los Alamos group for the determination of the equation of state of a large number of metals over a wide (and very high) pressure range (Walsh, et al., 1957). Due to the fact that the data are obtained by measuring the arrival time at two points, spaced roughly $\frac{1}{4}$ inch apart, details of the shock wave are missed. Developments in techniques to overcome these shortcomings have taken two forms. Both are directed at measuring particle velocity by measuring the free-surface velocity. The first (which was suggested by Davies (1948) and used to study elastic waves in rods) involves the measurement of the change in capacitance between a free surface and a second surface spaced at some distance from it. The problem, of course, is the development of suitable circuitry so that a good space resolution may be obtained where the total event time is of the order of 10^{-6} seconds. When a shock wave reaches a free surface, the surface reflectivity has been observed to decrease. Similarly, when a fast-moving free-surface strikes a transparent plastic material, the plastic will become opaque. These observations form the basis of two other photographic techniques for measuring shock-wave particle velocity. However, the present state of the art of measuring these parameters may be best described by the eagerness with which workers active in this field are searching for new ideas. As quantitative study of the structure of strong shock waves in solids promises to provide new and basic information on the behavior of solids, it would seem important to support development of techniques and instrumentation to study shock wave parameters and to encourage and support new ideas which might show promise in the difficult area of direct observation of pressure, density and structural changes that occur when a strong shock wave passes through a solid.

Attempts to observe the pressure profile of a shock pulse directly with a transducer have been frustrated by the difficulty of passing a strong shock pulse through such a transducer without extensive fracture. This difficulty seems to be inherent in ionic or covalent crystals in which the mechanical properties are extremely strain-rate dependent. There have been attempts to make direct radiographic observations of density changes in a material subjected

to a strong shock wave. Suitable flash X-ray tubes were developed both in the United States and Europe during the war. Edge effects are an inherent problem in this technique which would seem to limit its usefulness for quantitative work. The development of a flash X-ray diffraction technique would seem to be the answer to the edge-effect problem with the additional bonus that structural changes during the pressure-induced phase transitions could be studied. The problem of obtaining X-ray diffraction data in a fraction of a microsecond is a formidable one, and by most standards, would certainly be an expensive one. However, it would seem that a serious investigation of this technique is justified in view of the important results that would be forthcoming.

5.6. Structure and Property Changes

Some attention has been paid in the past to the changes in microstructure, and accompanying changes in properties, which often occur in metals subjected to a strong shock wave (Smith, 1958). For certain metals, principally copper and iron, interesting structural features have been observed metallographically. This work has been, and promises to be, important scientifically as a tool for studying shock waves. Metal-forming techniques that utilize high-explosives (or other means for impulsive loading) and a single die are rapidly becoming important technologically. The development of a suitable technology should be based on sound investigations of the structural and property changes that occur under these loading conditions. Work of this type should be based on the best possible quantitative description of the entire stress system and, at least initially, on simple metal specimens, well characterized as to structure and composition.

Impulsive loading techniques provide a powerful tool for studying the fracture of ductile materials in the absence of large deformation (Rinehart, et al., 1954). A compression wave will reflect from a free surface as a tension wave. Normally a net tension will be produced in the material just behind the free surface. If quantitative data is available on the pressure profile of the incident compression wave, it is possible to determine accurately the net tension behind the free surface as a function of the distance from the free surface. Observations of the position of fracture relative to the free surface provide quantitative information on the fracture stress. If the specimen of interest is inserted as a pellet in the front surface of a plate, it is possible to control the time of application of the stress and to estimate the rate of crack propagation. The stress distribution within the pulse is obtained by a modified

Hopkinson (1905) bar technique which involves measurement of the velocity of a number of pellets that are ejected from the free surface of the plate. Present techniques (which are good to about one part in a hundred) are adequate for stress measurements for fracture. However, if it were possible to develop a technique for measuring the velocity of a pellet (traveling at some 1,000-2,000 feet per second) to within one part in 10,000 it would then be possible to measure directly the energy required to initiate and propagate a fracture in a solid under these loading conditions.

To summarize, it is important to distinguish between experiments where the stress wave is long compared with the diameter of the specimen, i.e., the rod configuration, and experiments where the stress wave is short relative to the diameter of the specimen, i.e., in the plate configuration. Although the former is important in studies of the mechanics of stress waves in solids, it has limited usefulness in studying the behavior of materials. Investigations based on the plate configuration have been, and will continue to be, important for obtaining compressibility data at very high pressures, data on dynamic yielding, pressure-temperature phase diagrams, fracture data on ductile materials in the absence of large strains, etc. Improvements in technique and instrumentation for shock wave and particle velocity measurements are necessary for the detailed study of the pressure profile at the front of a strong shock wave. The development of techniques for the direct observation of pressure, volume, and crystal structure behind the front of the shock wave could be extremely fruitful.

5.7. References

- Bell, J. F., 1959, *J. Appl. Phys.*, **30**, 196.
Clark, D. S., and Wood, D. S., 1949, *Proceedings ASTM*, **49**, 717.
Davies, R. M., 1948, *Phil. Trans. Roy. Soc.*, **240A**, 375.
Hopkinson, B., 1905, *Proc. Roy. Soc.*, **74**.
Kolsky, H., 1953, *Stress Waves in Solids*, Clarendon Press, Oxford.
Minshall, S., 1955, *J. Appl. Phys.*, **26**, 463.
Rice, M. H., McQueen, R. G., and Walsh, J. M., 1958, *Solid State Physics*, **6**.
Rinehart, J. S., and Pearson, J. P., 1954, "Behavior of Metals Under Impulsive Loads," *Am. Soc. Metals*, Cleveland, Ohio.
Smith, C. S., 1958, *Trans. AIME*, **212**, 574.
Walsh, J. M., and Christian, R. H., 1955, *Phys. Rev.*, **97**, 1544.
———, Rice, M. H., McQueen, R. G., and Yarger, F. L., 1957, *Phys. Rev.*, **108**, 196,

6. HIGH MAGNETIC FIELDS

6.1. Introduction

During the last five years, very high magnetic fields have become a principal tool for solid state research. It is clear that the importance of such fields will continue to increase for many years to come.

Magnetic fields from the very small to the 20,000 to 40,000 gauss region have long played a crucial role in physics. Larger fields, larger than 40,000 gauss, are produced by large currents in air-core solenoids. Direct-current fields up to about 100,000 gauss are now available in very few laboratories. Pulsed fields of 1,000,000 gauss and more have been reported. It now appears that dc fields up to region of 250,000 gauss can be produced with a modest expenditure. Steady magnetic fields of this intensity can be tremendously useful in advancing our knowledge in solid state physics and in other areas of science.

The earliest high-field work with air-cored solenoids was done by Kapitza (1924) at Oxford from 1923-1934. He produced (and used for a variety of solid state and other experiments) fields of 100,000 gauss for short periods in a battery-powered magnet and 300,000 gauss in a pulsed magnet powered by short-circuited generators.

Bitter (1936a, 1936b, 1937, 1939) built a new type of solenoid magnet which with an input of 1.7 megawatts of dc power produced 100,000 gauss continuously. This facility is still in operation at M.I.T.

Since World War II, several large magnet installations have been established and several more are now being built or planned. With the exception of the one recently proposed by M.I.T., all of these post war magnet facilities are designed to produce fields no larger than 100,000 gauss. A number of pulsed magnet installations are in existence today. These have been used for experiments in solid state physics, for plasma physics and shock wave studies in ionized gases. The ability of pulsed magnets to provide fields from 250,000 to over a million gauss makes them useful for exploratory investigations. However, the short duration and rapid variation of magnetic field make their use difficult for some experiments and impossible for others. Well-controlled, continuous fields in the range from 40,000-250,000 gauss would be a unique and vitally important tool for solid state research.

6.2. Experimental Programs

Some of the general classes of experiments for which very high fields are necessary will be listed and briefly discussed here. An equally exciting prospect is the almost certain discovery of new phenomena when experiments are performed in this heretofore unexplored environment.

6.2.1. Cyclotron Resonance in Solids

Cyclotron resonance is a powerful tool for the determination of the electronic band structure of solids (Lax, 1958). In most materials, the short electron mean free times makes observation of cyclotron resonance possible only at millimeter wave or infrared frequencies. The corresponding magnetic fields needed for resonance are then well above those obtainable from conventional iron core electromagnets. Direct-current magnets are necessary since pulsed magnets are unsuitable in the far infrared.

6.2.2. Magneto-optic Effects

Magneto-optic effects in solids, including Faraday rotation, Zeeman splitting and oscillatory absorption effects, also provide important information on the electronic band structure (see Review articles, 1959). In most cases, these phenomena are observable only at high fields. In general, the higher the field, the greater the effect. In addition to the infrared region now beginning to be explored, the visible and the ultraviolet regions of the spectrum should be extensively studied.

6.2.3. Galvanomagnetic and Thermomagnetic Effects

These effects result from the interaction of electric and magnetic fields and thermal gradients. They provide information on the amount and nature of the electronic scattering processes taking place during electrical and thermal conduction. Much work has been done at low magnetic fields, but many phenomena can best be studied in very large fields.

6.2.4. Magnetic Resonance

Very high magnetic fields are needed to observe resonance phenomena in many ferromagnetic, ferrimagnetic and antiferromagnetic materials. Some of these resonances occur in the millimeter range and have been studied in pulsed magnetic fields (Foner 1957, 1959). Others occurring in the millimeter wave and infrared

regions, where there is limited source intensity and detector sensitivity, must be observed with dc magnetic fields.

6.2.5. De Haas-Van Alphen Effect

The oscillatory magnetic susceptibility or De Haas-van Alphen effect is another powerful tool for determining band structure of solids. It has been particularly fruitful in the case of metals and semimetals (Schoenberg, 1939, 1952, 1953). Here again this effect is evident only at high fields and in general, the higher the field, the better.

6.2.6. Low Temperature Research

Large dc magnetic fields in relatively large volumes will facilitate the attainment of very low temperatures by adiabatic demagnetization and make possible resonance and other experiments at temperatures below 1°K.

6.2.7. Magneto-Acoustic Resonance and Oscillatory Effects

These recently discovered effects provide information on both the scattering and band structure of metals (Bömmel, 1955; Pipard, 1957; Morse, et al., 1958; Reneker, 1958). High magnetic fields will be important in these studies.

6.2.8. Millimeter and Submillimeter Wave Generation

Very high fields set the cyclotron frequency in many materials into the millimeter and submillimeter range. It seems likely that this phenomenon could be used to provide a coherent source of power in this difficult spectral region. Other practical applications undoubtedly await fuller exploration of the properties of matter in high magnetic fields.

6.2.9. Mechanical Properties of Solids

In addition to the solid state experiments mentioned above, high fields can be used to study mechanical properties of materials. As mentioned before, soft conductors as copper, silver, etc., flow under magnetic stress imposed by fields in the region of 300–500 kilogauss. Magnetostriction of magnetic materials has not been explored in the 100 kilogauss range or above. Finally, one could magnetically induce shock waves in solids by large pulsed fields.

Many other uses of high magnetic fields in areas not directly related to solid state physics are evident. Some of these areas are plasma dynamics, nuclear alignment, cosmic ray physics, and Zeeman spectroscopy.

6.3. Present Possibilities

The present state of the art in magnet design is such that one or several facilities could be built to achieve fields as high as 250,000 gauss in a coil approximately one inch in diameter. This would require a power supply of about 8 megawatts continuous rating. The power source would consist of four identical 2 megawatt units providing 10,000 amperes at 200 volts each. These units could be connected in any series or parallel combination, or could be operated independently. Thus, four separate magnets dissipating up to 2 megawatts each could be operated simultaneously.

6.4. Future Possibilities

6.4.1. Large Water-Cooled Coils

It is conceivable that fields of the order of 500,000 gauss in larger internal diameters of the order of approximately 6 inches, which may be required to solve the cooling problem, can be achieved with present techniques. From the fundamental power relation, if the resistivity of the copper is assumed to be the same and the filling ratio of conductor to insulation also remains constant, then $P \sim dH^2$ where d is the internal diameter and H is the peak field. If we compare a 1 in. 250,000 gauss magnet with a 6 in. 500,000 gauss magnet, the power requirement goes up by a factor of 25, and would be about 200 megawatts. At the present time, this exceeds the capability of any existing power supplies.

6.4.2. Transient or Intermittent Operation (Flywheels)

Both the Stellarator and the proposed Cambridge Magnet Facility will be equipped with motor generator sets which have flywheels. This arrangement will allow transient operation of the generators at approximately four times the rated power for periods of about 3 seconds. With appropriate control of the exciter, the wave-shape of the current input to the coils can be maintained flat for this period. Such an operation will allow the performance of a selected number of solid state and other experiments listed under section 6.2 with fields between 300,000 and 500,000 gauss.

The increased power into the coils could be handled in one of two possible ways. The cooling system of the coil can be increased to handle the full peak power, and the coils would be designed

accordingly. The alternative would be to make the coils large and allow them to be heated up on a transient basis and to cool down between the intermittent pulses. In any event, the operation of coils under these higher fields and transient conditions would require improved mechanical design, since copper begins to flow under the pressure set up between adjacent conducting elements by fields in excess of about 300,000 gauss.

6.4.3. Low Temperature Operation

The possibilities of operating dc magnets at low temperatures is attractive since it is possible to operate coils at a reduced power level. Thus, copper at liquid hydrogen temperatures has a resistivity of ~ 0.01 of that at room temperature, thus requiring much less power. This approach has been carried out by the Los Alamos group (Laquer 1956, 1957), and requires a good deal more research and development before it becomes competitive with the water-cooled system. Another possibility, utilizing sodium as the conductor imbedded in stainless and operating at $\sim 10^\circ\text{K}$ with helium cooling, is being considered by the Livermore group. This is still in its early stages, but may have possibilities for high field operation. This approach, although attractive from theoretical considerations of the electrical properties of pure sodium, introduces many practical difficulties which will not be overcome without considerable research and development on large-scale closed systems, low temperature systems, as well as on high-current, low-voltage generators.

6.5. Pulse Magnets

The achievement of high fields by the pulsed system using stored energy, i.e., capacitive, inductive or battery discharge, should be pursued along two principal directions (Furth, et al., 1956, 1957; Foner, et al., 1956). First, high fields of the order of 10^6 gauss and $10\ \mu\text{second}$ duration in relatively small volume should be designed so as to survive repeated pulses. The flux concentrator principle which relies on skin depth effects seems the most promising approach. The second direction of fruitful exploration is that of developing larger-volume, long-duration, pulsed systems in the millisecond range. A goal to aim for in this category may be 10^6 gauss in a volume of approximately 10–20 cubic inches. It is estimated that such a system would require $\sim 10^7$ joules of stored energy and would cost approximately two million dollars. The main problems involve proper coil design and switching de-

vices. Low temperature operation and the use of pulse transformers may be necessary if it proves to be difficult to remain in the oscillatory domain. In addition, such an arrangement would permit the use of ignitrons instead of air gaps. This may be desirable.

Such fields with a time constant in the 1-100 millisecond range would permit the extension of resonance experiments to additional materials such as semimetals, antiferromagnets with high Curie temperature, and galvanomagnetic measurements in semiconductors to still higher fields.

6.6. Superconducting Magnets

Within the last two years there has been a growing interest in superconducting magnets for a number of possible applications. The first of these has been their utilization at moderate fields in connection with masers. For this purpose, Stan Autler of Lincoln Laboratory in 1959 developed a niobium coil capable of providing about 10,000 gauss. Subsequently, using niobium and a high permeability steel, he developed an electromagnet capable of achieving 25,000 gauss in small volumes. The breakthrough came late in 1960 when a group at the Bell Telephone Laboratories developed a technique for making a type of niobium-tin wire which remained superconducting while carrying an appreciable current in an external magnetic field of 88,000 gauss. More recently, other alloys such as niobium-zirconium and technetium-molybdenum have been found to retain their superconducting properties in relatively high fields. At this writing, a magnet constructed with niobium-zirconium has produced fields as high as 43,000 gauss and further development work is being vigorously pursued in many laboratories.

It seems clear that superconducting magnets will be a useful tool for materials research. However, they will complement rather than replace many types of electromagnets in existence today or those aimed at extremely high fields like the pulse magnets and the large water-cooled magnets being designed at M.I.T. It is of course conceivable that perhaps five to ten years from now in some respects the superconducting magnets may become competitive with the latter above the 100,000-gauss range. However, it is the consensus of opinion that a great many technological problems will have to be solved for this to occur. In the meantime, in the region of 50,000 to 100,000 gauss, where large volume magnets are required for such things as fusion, magneto-hydrodynamics and

other applications, this looks like the best approach, although a difficult one. Consequently, it is not surprising that Oak Ridge, the other AEC Laboratories, and commercial organizations are applying a great deal of effort to achieve this objective. Such magnets, when developed and used in conjunction with the water-cooled coils which have greater mechanical strength to withstand very high fields, will permit the latter, when enclosed in a large superconducting magnet, to increase the field by an additional 100,000 gauss. Thus the combination will provide 350,000 gauss or more instead of the 250,000 gauss presently planned.

6.7. References

- Bitter, F., 1936a, *Rev. Sci. Instr.*, **7**, 482.
———, 1936b, *Rev. Sci. Instr.*, **7**, 479.
———, 1937, *Rev. Sci. Instr.*, **8**, 318.
———, 1939, *Rev. Sci. Instr.*, **10**, 373.
Bömmel, H. E., 1955, *Phys. Rev.*, **100**, 758.
Foner, S., and Kolm, H. H., 1956, *Rev. Sci. Instr.*, **27**, 547.
———, 1957, *Phys. Rev.*, **107**, 683-685.
———, 1959, *Jour. Phys. et le Rad.*, **20**, 336.
Furth, H. P., and Waniek, R. W., 1956, *Rev. Sci. Instr.*, **27**, 195.
———, Levine, M. A., and Waniek, R. W., 1957, *Rev. Sci. Instr.*, **28**, 949.
Kapitza, P., 1924, *Proc. Roy. Soc. (London)*, **A105**, 691.
Laquer, H. L., 1956, *Proc. Inst. Soc. Am.*, **56**, 16.
———, 1957, *Rev. Sci. Instr.*, **28**, 875.
Lax, B., 1958, *Rev. Mod. Phys.*, **30**, 122-154.
Morse, R. W., Bohn, H. V., and Gavenda, J. D., 1958, *Phys. Rev.*, **109**, 1394.
Pippard, A. B., 1957, *Phil. Mag.*, **2**, 1147.
Reneker, D. H., 1958, *Phys. Rev. Ltrs.*, **1**, 440.
Schoenberg, D., 1939, *Proc. Roy. Soc. (London)*, **A170**, 341.
———, 1952, *Proc. Roy. Soc. (London)*, **A245**, 891.
———, 1953, *Physica*, **19**, 791.

6.8. Review Articles

1959, *J. Phys. Chem. Solids*, Vol 8, Articles by:

- E. F. Gross
H. Y. Fan and P. Fisher
E. Burstein, G. S. Picus, R. F. Wallis, and F. Blatt
B. Lax, L. Roth, and S. Zwerdling
R. F. Wallis and H. J. Bowlden
W. S. Boyle
T. S. Moss, S. D. Smith, and G. W. Taylor
R. J. Elliott and R. London
S. Zwerdling, L. Roth, and B. Lax

7. HIGH FREQUENCY ULTRASONICS

7.1. Introduction

The propagation behavior of high frequency elastic stress waves in solids provides an effective and a sensitive means for the study of quite a few fundamental physical properties of solid materials. The propagation behavior of these stress waves is characterized by the attenuation (or energy loss factor) and the velocity of the wave (assuming a plane wave). The methods are particularly sensitive to changes in the nature and number of defects present as well as to the manner in which the defects are distributed in the material. There are interactions between dislocations and ultrasonic waves—interactions which permit the study of dislocation dynamics, (Granato and Lücke, 1956) (Hikata, et al., 1956) particularly as the dynamics are influenced by interaction with point defects. There are interactions between electrons and ultrasonic waves (Morse, et al., 1958, 1959) at low temperatures which produce energy losses and velocity changes. These effects are observed by studying attenuation as a function of the magnitude and orientation of an applied magnetic field. There are interactions between ultrasonic waves and energy levels arising as the result of the orientation of nuclei in the field of the crystal lattice; nuclear spin transformations in solids have been induced by ultrasonic waves of proper frequency. The ultrasonic wave is coupled to the nuclear spin through the interaction of the quadrupole moment of the nucleus with an electric field gradient produced by the elastic stress waves in altering the positions of the charges nearest the nucleus in question.

There are thermoelastic losses (Lücke, 1956) connected with the passage of high frequency ultrasonic (compressional) waves through a solid. The path of the heat flow is of the order of a half wavelength which, in the high frequency (megacycle) range, is less than about one mm and may be (at present) as small as one micron or less. Thermoelastic phonon loss effects are particularly interesting in the kilomegacycle frequency region, where the phonon-phonon relaxation time, τ , satisfies $\omega\tau = 1$ at a temperature in the region $10^\circ - 100^\circ\text{K}$ for many crystals.

There are defects which scatter ultrasonic waves (Ying and Truell, 1956; Einspruch and Truell, 1960). In principle all lattice defects will scatter, and for a single scatterer, the amount depends mainly (although not entirely) on the size of defect and the frequency of the ultrasonic radiation used. The amount of scattering also depends, of course, on the density of defects. A type of defect

that is of interest in connection with scattering is the damage region produced by fast neutrons in certain kinds of solid materials. Another specific type of defect is a precipitate. Any region that is elastically different from the lattice matrix constitutes a scatterer.

There are interactions between ultrasonic waves and magnetic domain walls in ferromagnetic materials. The attenuation changes with magnetic field in nickel, for example, (Levy and Truell, 1953; deKlerk, 1959; West, 1957) may be very large depending on the propagation direction in the crystal. The velocity changes are small but measureable.

There are interesting "physical optics" applications of stress waves; one of these effects results in ultrasonic double refraction produced by anisotropic strains (Waterman and Teutonico, 1957). Anisotropies introduced into a lattice by defects having a directional character have been studied by means of attenuation and velocity measurements (Truell, et al., 1957). Another effect is that of internal conical refraction resulting from the inherent elastic anisotropy of single crystals when one propagates shear waves in the [111] direction in cubic single crystals.

Other effects were predicted and have now been demonstrated. The population of energy levels of paramagnetic ions may be altered by ultrasonic energy and this has been detected by spin resonance methods (Kittel, 1958; Jacobsen, 1959a, 1959b). Such experiments have determined the feasibility of the methods as detectors and possibly as generators of microwave phonons. Masers using ultrasonic pumping appear to be feasible, and if so, might have advantages of lower noise level. It may be possible to observe by acoustic methods the "drag effects" of phonons in solids. With the availability of pulsed sound at thousands of megacycles it seems to be possible to observe directly the superconducting energy gap (temperature dependent) which appears in the density of states function at the Fermi energy.

It seems clear that higher frequency and improved techniques are certain to contribute new information in the study of the physical properties of materials. Some of the above topics, such as the nuclear effects, and magnetic domain effects, have been omitted from further discussion here because, while there is much work that can be done in these areas, it is at present not clear that higher frequencies will contribute new information. On the other hand, it seems rather clear that in such areas as dislocation dynamics, thermoelasticity, the electron properties of solids, scattering by lattice defects, and electron spin resonance effects, the

higher frequencies will contribute significantly to better understanding of solid materials.

7.2. Attenuation Mechanisms

7.2.1. Dislocation Damping

In connection with the losses accompanying the vibrational motion of dislocations under the influence of a rapidly alternating stress field, there seems to be a fairly satisfactory mathematical picture of the damped, driven, oscillating system (for a dislocation), but there is no satisfactory physical theory for the specific damping coefficient necessary to the mathematical expression for this damped vibrating system. It is not even established whether the specific loss mechanism is basically electrical or mechanical in nature; it is possible that a vibrating dislocation may have electromagnetic radiation losses. Higher frequency work in the range between 100 mc/sec and 2,000 mc/sec should show whether the present dislocation damping theory is supported in its prediction that a resonance of the dislocation should occur in this frequency range. The measurements on single crystals of pure metals and semiconductors, using both compressional and shear waves, should enable the losses due to dislocation damping to be separated from those due to thermoelasticity, at least in some cases.

7.2.2. Thermoelastic Effects (Umklapp Processes)

The attenuation of compressional sound waves due to thermoelastic heat flow (Lücke, 1956) is a relaxation mechanism, and so depends strongly on the frequency, ω , of the sound waves and the relaxation time, τ , of the elastic after-effect. In the kilomegacycle frequency region, $\omega\tau \ll 1$ at room temperature, but $\omega\tau \gg 1$ at low temperatures. The thermoelastic losses due to heat flow from compressed to expanded regions of a solid during the passage of the sound wave are small compared with the losses of another relaxation process with a similar frequency behavior, namely, those which arise from perturbing the equilibrium distribution of the phonons.

The latter relaxation process is the return of the thermal phonon energies from the nonequilibrium distribution, caused by the passage of the sound wave, to the equilibrium distribution characteristic of a particular lattice at a particular temperature. It is assumed that in the lattice there are collisions among thermal phonons of the type which Peierls called the Umklapp process.

Such collisions of the Umklapp type are evidence of an anharmonic coupling between the lattice vibrations and the sound vibrations and permit a change in the direction of energy flow, thereby involving different branches of the lattice vibration spectrum, which may be at different temperatures.

The Umklapp process collisions should disappear at very low temperatures because the rate at which such processes occur diminishes with temperature, T , as $\exp(-y\theta/T)$ where y is a constant of order 1 and θ is the Debye temperature.

Measurements at high frequencies with compressional waves in quartz (Bömmel and Dransfeld, 1959) and germanium (Dobbs, et al., 1959) show that the main part of the measured attenuation has disappeared by the time the temperature is lowered to 20°K in a manner that strongly suggests the dying out of Umklapp processes. In the case of germanium the shear wave attenuation has been found to be considerably smaller than the compressional wave attenuation between 20°K and room temperature, thereby confirming the fact that the attenuation is largely of a thermoelastic type in this case.

7.2.3. Phonon-Phonon Interactions

The phonon relaxation time τ satisfies $\omega\tau \gg 1$ at high enough frequencies and at low enough temperatures. In this limit, the acoustic energy may be considered as acoustic phonons, and the interaction between acoustic phonons and thermal phonons may, in principle, be studied (Simons, 1957). Under suitable experimental conditions, a direct verification of basic phonon-phonon interaction theory should be possible. In this field, acoustic experiments can yield results over a narrow band of phonon wavelengths, whereas thermal conductivity measurements necessarily involve averaging over many wavelengths. An additional advantage of the acoustic absorption method is that a narrow pencil beam of energy can be propagated and the size of the specimen, which effectively limits the measurement of lattice thermal conductivity below about 10°K, is not a limitation. Moreover, the power of the acoustic source can be varied, thereby varying the acoustic phonon intensity while maintaining the sample at a constant low temperature, since narrow pulses of ultrasound are available. No measurements of these interactions have been reported thus far.

7.2.4. Scattering Processes

The scattering problems become more interesting and important at the higher frequencies. The scattering cross-section increases as ν^n where n lies between one and four. The range of fourth power frequency dependence defines the range of Rayleigh scattering. Defects of smaller sizes and in lower concentrations can be studied better at higher frequencies.

Information about sizes of defects, their concentration, and the changes in these quantities with time and other parameters are available under proper conditions. Attenuation as a function of frequency can yield information about whether the scattering is Rayleigh or non-Rayleigh, hence, information of the sizes of scatterers relative to the wavelength used.

7.2.5. Electron-phonon Interaction (Metals)

Experimental evidence of interaction between ultrasonic stress waves and electrons in metals has been supplied by Bömmel (1959) and Morse, et al. (1959b) in the neighborhood of the superconducting transition. The general problem of the interaction of electrons with elastic waves (scattering) is as old (at least) as the theory of lattice vibration and specific heats. The relatively recent developments have been concerned with recognition that while electron mean free paths are normally shorter than ultrasonic wavelengths at ordinary temperatures, this situation may be reversed in very pure metals at low temperatures; i.e., electron mean free paths may be longer than ultrasonic wavelengths. Many discussions of the theory of the electron-phonon problem have appeared and a review of the work can be found in Morse and Gavenda (1959b) and Cohen, et al. (to be published). The use of the magnetic field dependence of ultrasonic attenuation for studying Fermi surfaces in metals has been pursued by Morse, et al. (1958a, 1958b, 1959) and Olsen and Morse (1959) in copper, and by Reneker (1958) in bismuth, and by Neighbors and Alers (1959) in copper.

Attenuation studies in superconductors together with the Bardeen, Cooper, and Schrieffer theory of superconductivity have shown excellent agreement between theory and experiment on indium and tin. The theory makes rather clear and explicit predictions about the variation of the electron-phonon behavior in the superconducting state. The temperature-dependent energy gap $[2\Delta(T)]$ which appears at the Fermi energy in the density of states $\rho(E)$ has been studied by means of attenuation measure-

ments in the normal and superconducting states. The agreement between theory and experiment for the evaluation of $2 \alpha(T) = 3.50 \text{ kT}_c$ at absolute zero is amazingly good.

7.3. Equipment and Techniques

Attenuation and velocity measurements at the higher megacycle frequencies are at present accomplished mainly by use of pulse echo techniques (deKlerk, 1959; Roderick, et al., 1952; Chick, et al., 1960; Bosworth, et al., 1951; Huntington, et al., 1947; Espinola, et al., 1958; Bömmel, 1958). A pulse of ultrasonic energy is obtained by applying an electric field, also pulsed, across a piezoelectric material which is usually, although not necessarily, quartz. If the quartz is properly oriented with respect to the direction of the applied electric field, a shear wave or a compressional wave may be generated. The wave so generated is next coupled into the medium of interest through a bond between the piezoelectric transducer and the medium. The sample medium has two flat parallel faces separated by a distance depending on the pulse length and the velocity in the sample medium. The pulse is sent through the medium under study and either reflected back to the same transducer or received by a second transducer bonded to the opposite face of the medium under study. From a single pulse introduced into the sample there are many multiple echoes arising from having made one, two, three or many round trips in the sample. Each time the pulse returns to the transducer-sample interface a potential difference is generated across the transducer, and this signal is received, amplified, and applied to an oscilloscope. On the oscilloscope is observed a train of attenuating pulses each one of which has made one more round trip in the sample than the preceding pulse.

The attenuation value is obtained either by use of a calibrated exponential curve presented on alternate sweeps on the oscilloscope, and which can be fitted to the echo pattern, or by means of a comparison pulse, which can be moved to be adjacent to any echo in the pattern and adjusted in height by means of a calibrated attenuation box (Roderick and Truell, 1952; Chick, et al., 1960). The time interval between pulses measures the velocity when the sample thickness is known.

As the frequency of the operation is increased, the parallelism of the opposite faces of the sample (and the transducer) becomes more and more important. At thousands of megacycles the question of parallel faces is a critical matter. The quality of the bond between transducer and bond also becomes more and more

important as the frequency increases. In fact the bond is one of the most critical features, and even with the best bonds that can be obtained, it is necessary to measure the bond loss as a function of the frequency, if absolute attenuation measurements are sought. The bond loss increases with frequency, and at sufficiently high frequencies the bond loss may be larger than the material attenuation.

The highest frequencies used thus far (up to 25,000 mc/sec and at low temperatures) have used a quartz rod as the medium where one end of the rod was the transmitter and the other was the receiver. One end of the rod was placed in the gap of a cavity resonator where the oscillating electric field "drove" the end face of the quartz rod as one face of a transducer. The other end of the rod, placed in another resonator gap, acted as a receiver. It is also possible to use one end of the rod as a source and a receiver but this requires a T.R. box or the equivalent which complicates the system somewhat.

There are several disadvantages of the quartz rod and cavity method. First of all, the wave propagates as a rod or guided wave mode and the sidewalls cause a complicated wavefront which makes attenuation and velocity measurement difficult. Second, at the highest frequencies (5,000 to 25,000 mc/sec) one cannot tune over a frequency range in the way one can up to a few thousand megacycles/sec, so that a separate measurement system is necessary for each point in the frequency range. Thirdly, the measurements at the highest frequencies require working at low temperatures, usually below about 50°K. Making a bond that will transmit energy above 5,000 mc/sec is very difficult although absorption measurements have been made in a thin layer of indium bonded between two quartz rods. In general there is difficulty when it is necessary to bond a transducer to a sample for work at the lowest temperatures. The problem of making such a bond between a transducer and a material of very dissimilar thermal expansion properties is one of the most troublesome parts of this type of work. Frequently the sample is deformed in the course of lowering the temperature; sometimes a bond cannot be maintained at all at very low temperatures. In some cases it is possible to lower the temperature part way, make the necessary bond, and then lower the temperature further.

The techniques involved in making really accurate velocity measurements (0.1% or better) are difficult. Velocity measurements made under fixed conditions of frequency and state of the material are relatively less difficult than when the velocity is

required as a function of frequency over an appreciable frequency range for studying dispersion, or when velocity measurements are needed as a function of deformation or irradiation. In these latter cases, the attenuation, and hence the pulse shape, may change during the process. A fundamental limitation in velocity measurements is that these measurements must be made on a pulsed r.f. wave that is delayed by velocity changes but which is not distorted under any conditions.

All of the absolute velocity measurement methods in use at the present time have difficulties of a fundamental nature. A really detailed analysis of each method is necessary to show the difficulties. In most cases the accuracy is difficult to determine. In some cases the reproducibility of the measurements is much better than the accuracy to be expected.

Sensitive and accurate velocity measurements are needed with attenuation measurements to identify and study loss mechanisms. In addition there is a strong need for determining elastic constants with sufficient accuracy to be able to decide whether the differences to be noted in the literature are really differences in the material or the result of measurement difficulties. The elastic constants are concerned with the binding forces, thermal properties, dislocation effects and other imperfection effects in solid materials. It now appears that measurable velocity changes also arise from electron-phonon interaction at low temperatures (Gibbons and Renton, 1959).

7.4. References

- Bömmel, H. E., and Dransfeld, K., 1958, *Phys. Rev. Ltrs.*, **1**, 234; 1959, *ibid*, **2**, 298.
- Bosworth, R. M., Mason, W. P., and McSkimin, H. J., 1951, *Bell Sys. Tech. J.*, **970**.
- Chick, B., Anderson, G., and Truell, R., 1960, *J. Acoust. Soc. Am.*, **32**, 186.
- Cohen, M. H., Harrison, M. J., and Harrison, W. A., *Phys. Rev.*, to be published.
- deKlerk, J., 1959, *Proc. Phys. Soc. (London)*, **73**, 337.
- Dobbs, R., Chick, B., and Truell, R., 1959, *Phys. Rev. Ltrs.*, **3**, 332.
- Einspruch, N. G., and Truell, R., 1960, *J. Acoust. Soc. Am.*, **32**, 214.
- Epinola, E. P., and Waterman, P. G., 1958, *J. Appl. Phys.*, **29**, 718.
- Gibbons, D. F., and Renton, C. A., 1959, *Phys. Rev.*, **114**, 1275.
- Granato, A., and Lücke, K., 1956, *J. Appl. Phys.*, **27**, 583.
- Hikata, A., Truell, R., Granato, A., Chick, B., and Lücke, K., 1956, *J. Appl. Phys.*, **27**, 396.
- Huntington, H. B., 1947, *Phys. Rev.*, **72**, 321.
- Jacobsen, E. H., 1959a, *Phys. Rev. Ltrs.*, **3**, 83; 1959b, *ibid*, **2**, 249.
- Kittel, C., 1958, *Phys. Rev. Ltrs.*, **1**, 5.
- Levy, S., and Truell, R., 1953, *Rev. Mod. Phys.*, **25**, 140.
- Lücke, K., 1956, *J. Appl. Phys.*, **27**, 1433.

- Morse, R. W., Bohm, H. V., and Gavenda, J. D., 1958a, *Phys. Rev.*, **109**, 1894; 1958b, *Bull. Am. Phys. Soc.*, **3**, No. II, 44.
- , and Gavenda, J. D., 1959a, *Phys. Rev. Ltrs.*, **3**, 250; 1959b, in *Progress in Cryogenics*, Vol. 1 (Heywood and Co., Ltd.).
- Neighbors, J. R., and Alers, G. A., 1959, *Phys. Rev. Ltrs.*, **3**, 265.
- Olsen, T., and Morse, R. W., 1959, *Bull. Am. Phys. Soc.*, **3**, No. II, 167.
- Reneker, D. H., 1958, *Phys. Rev. Ltrs.*, **1**, 440.
- Roderick, R. L., and Truell, R., 1952, *J. Appl. Phys.*, **23**, 267.
- Simons, S., 1957, *Proc. Camb. Phil. Soc.*, **53**, 702.
- Truell, R., Teutonico, L. J., and Levy, P. W., 1957, *Phys. Rev.*, **105**, 1723.
- Waterman, P. C., and Teutonico, L. J., 1957, *J. Appl. Phys.*, **28**, 266.
- West, F., 1957, private communication.
- Ying, C. F., and Truell, R., 1956, *J. Appl. Phys.*, **27**, 1086.

8. ANALYTICAL TECHNIQUES

8.1. Introduction

The importance of chemical analysis in materials research is obvious. Analyses for major constituents usually pose no great problems, although new methods and procedures will need to be developed in order to analyze newer materials or to obtain higher precision. Analyses for trace or relatively small amounts, on the other hand, frequently pose problems for which there are no adequate solutions up to the present time. The following discussion will concentrate for the most part on the status of analytical techniques for such low concentrations and will attempt to point out the promising paths for additional investigation.

8.2. Chemical and Spectrographic Methods for Small Concentrations

The direct application, without prior concentration of impurities, of standard chemical and spectrographic methods will continue to be useful down to the few ppm range and in special cases somewhat lower. However, further developments of these techniques appear to have little potential for general application at levels much below 1 ppm. On the other hand, prior concentration of the trace or impurity elements followed by more or less standard chemical and spectrographic analytical procedures has great usefulness. Concentration has been effected by a variety of means. One of the most attractive and useful procedures is that of selective adsorption and elution of the trace elements on resin columns (Brody, et al., 1958; Faris, 1958; Buchanan, et al., 1958), by which the major constituents (matrix) are effectively removed. Prior concentration has also been done very effectively

by means of solvent extraction and by volatilization of the matrix. Relatively large samples can be accommodated and the concentrated trace elements are then analyzed in the usual ways for the individual components. Zone melting (Davis and Chalmers, 1958) has been suggested as a means of concentration of trace elements. Other versions of the selective adsorption and elution methods, which in a more general way are termed chromatographic techniques, are also useful (Strain, 1958). These techniques appear to have tremendous potentiality and warrant considerable development work. The applicability is limited in a general way to those systems where the matrix elements and the trace elements have considerably different adsorption properties. However, the limitations can certainly be decreased greatly by further studies and by development of new adsorption and elution systems, as, for example, the possible use of non-aqueous hydrogen fluoride rather than resins. For complete exploitation of these techniques, it will be necessary to develop purer resins and to employ much better cleanliness techniques. The analyst will be required to work in special areas with extreme care given to contamination problems. Presently attained sensitivities are in the several ppb range, but where the sample size is not limiting, there appears to be no immediate lower level limitation.

8.3. Physical Methods for Small Concentration Analysis

Vacuum fusion techniques for carbon, nitrogen, oxygen and hydrogen are more or less routine in the 5 to 1 ppm range. The volatile components are frequently measured by manometric techniques. Specialized methods for measurements of the gases evolved, for example, the quenching of fluorescence by trace levels of oxygen or the measurement of magnetic susceptibility of oxygen, may aid to push the sensitivity down somewhat. However, adsorption of the gases which have been evolved on the walls of the vacuum envelope and other similar problems appear to pose great difficulties, and major increases in sensitivity by these techniques appear somewhat doubtful. On the other hand, the method, at the level of about 1 ppm is potentially useful for other elements, and a great deal more development work is called for to obtain adequate procedures for sulfur, selenium, phosphorus, silicon, and other elements which may be obtained as volatile components in vacuum fusion methods.

Isotopic dilution and other mass spectrographic methods have been found to be quite useful for very specialized problems. For example, such a method has proven very reliable for the analysis

of hydrogen in metals at the 1 ppm level (Holt, 1959). Such methods using separated stable isotopes have a rather wide applicability, and a great deal more work should be encouraged to further develop these procedures. On the other hand, the concentration sensitivities are modest, and applications in the level below 0.1 ppm are quite limited. On the other hand, further development of methods of ionization could make the mass spectrographic technique the most important single procedure for general analytical use at low levels. It is particularly useful where exceedingly small sample sizes are involved. A good summary of the mass spectrographic possibilities is given by Inghram (1959).

Various X-ray methods of analysis are frequently useful (Liebhavsky, 1958) but the impurity concentration sensitivity is not high, seldom better than about 10 ppm. With heavy scatterers in a light element matrix, X-ray fluorescence methods may attain 1 ppm sensitivity. Monochromatic X-ray absorption techniques can also reach a sensitivity of a few ppm in optimum cases. Development of more sensitive detectors and better ways of discriminating against Compton scattering are called for here. In certain cases, X-ray diffraction techniques can be useful in determining deviations from stoichiometry, or at least changes in the deviation from stoichiometric proportions. For example, it is possible to detect the deviation from stoichiometry in UO_2 corresponding to $\text{UO}_{2.0025}$, but the sensitivity is usually much inferior to this. In most cases the sensitivity is not better than that readily attainable by standard gravimetric procedures.

There is, of course, great and increasing interest in materials which do have small deviations in composition from stoichiometric proportions. Measurement of such deviations will require improved procedures of standard gravimetric analysis. This, in turn, will, in many cases, require new and improved values for the chemical weights of elements. Precision density determinations (say, by flotation techniques) of well-defined crystalline materials coupled with precision lattice parameter determinations may be the best approach to many of these problems.

Activation analysis appears to be one of the most promising of all techniques for rather general applicability in the ppb range. Although a good deal of research and development is underway, particularly at the Oak Ridge National Laboratory (Leddicotte, et al., 1959) and at the University of Michigan (Meinke, 1959), it appears that far too little effort is being put in this area considering its potential usefulness. Activation procedures should be developed using not only slow and fast neutron activation, but also

accelerated ion bombardments and perhaps even gamma and coulomb excitation. For example, the most sensitive method known for analysis of carbon is by means of the $C^{13} (d, n)N^{13}$ reaction. The present availability of pulsed electron linear accelerators at rather modest costs of about \$300,000 should make it relatively easy for many universities to acquire them and expand their analytical chemical research to embrace activation analysis studies. Such generators give a relatively high fast-neutron flux which would have advantages for the production of radioactive species by $(n, 2n)$, (n, α) , and (n, p) interactions. The machines would be particularly useful for the production of the short-lived activities which give the maximum sensitivity to the activation analysis method. Such activations followed by multichannel gamma ray spectrometry hold considerable promise for rapid and easy analysis at low levels. Methods of this type have been used by geochemists to attain analyses of phenomenally low concentrations (Hamaguchi, et al., 1957).

Other physical methods which show great promise for special systems include the electrochemical techniques. In the case of the anodic stripping techniques, the applicability is limited to impurities which can be reduced to the metallic state by electrolysis and thus deposited on the electrodes (Nikelly and Cooke, 1957). The sensitivity is in the 1 to 1/10 ppb range, and there is the advantage that no chemical reagents are required, which materially decreases the contamination. This technique should be particularly useful for analysis of depositable impurities in molten alkali halides.

Another electrochemical technique which is limited to either reducible or oxidizable impurities involves square wave polarography (Rosie and Cooke, 1955). Here, the voltage-current characteristics are determined in the stirred solutions and the impurity concentrations deduced. The sensitivity is limited approximately to 100 ppb—10 ppb. Further development of both of these electrochemical techniques appears very desirable, especially the development of molten salt systems in which one can dissolve the analytical sample for analysis. Chrono-potentiometry, as described by Laitinen (1959), also offers considerable promise for use in trace analysis.

There are many specialized techniques useful for specific materials; for example, the measurements of electrical properties such as conductivities and Hall coefficients of semiconductor materials. Also, the residual resistance of well annealed metals at low temperatures is useful as a relative index of "electronic"

purity without revealing the nature of the impurity. The method must be used with great caution since lattice defects are also effective scattering centers and are easily introduced by mechanical working or by quenching. Also, many of the common gases, H_2 , O_2 , N_2 , etc., dissolve in metals easily and may produce resistivity effects. Blewitt (private communication) has found that the residual resistivity of copper is greatly dependent on the annealing atmosphere. Ideally about 1 percent of impurities doubles the room temperature resistance of copper; thus, a 1 percent change in the liquid helium resistivity of copper corresponds roughly to about 1 ppm of added impurity (or point defects).

In the case of the alkali halide crystals, the electron trapping efficiency and the radiation coloration sensitivity is certainly a function of the impurity concentration. Unfortunately, in most of these cases one must know the level of concentration and the specific impurities involved in order to standardize the specialized technique. Other specialized techniques, for example the use of magnetic susceptibility measurements and magnetic resonance measurements, also have their applications.

Routinely attainable sensitivities for electron spin centers are in the range of 10^{12} spins for resonance of 1 gauss half width with 5×10^{11} being attained under optimum conditions (Smaller, private communication). Another factor of ten in sensitivity would appear to be about the lower limit since the theoretical signal/noise ratio quickly goes towards unity (Feher, 1957).

In certain cases the catalytic activity of impurities at surfaces may be an extremely sensitive measure of the impurity concentration. Similarly, the thermionic emission and field emission properties are very sensitive to the impurities. Any number of other impurity-sensitive properties could be mentioned, and most of them have had and will have application for analysis. However, these techniques appear to be a sort of stopgap until better standard analytical procedures can be developed. Review 1 is a useful general reference in this analytical area.

The status of analytical procedures and the importance in materials research of having a rather complete description of the material from the chemical point of view indicates that considerably more effort should be underway in this country on the further development of analytical techniques for low-level analysis. Many of our major research establishments appear to have adequate research leadership to guide this further work. A modest expansion in funds would make it possible for many of these organizations to add a number of supporting workers to their

existing staffs and some comparatively inexpensive equipment to meet this need for increased analytical research and development.

8.4. References

- Brody, J. K., Faris, J. P., and Buchanan, R. F., 1958, *Anal. Chem.*, **30**, 12, 1909-12.
- Buchanan, R. F., Faris, J. P., Orlandini, K. A., and Hughes, J. P., 1958, "The Analytical Application of the Nitric Acid Anion Exchange System to Plutonium-Fission and Plutonium-Binary Alloys Used in Metallurgical Studies," TID-7560, Technical Information Service Extension, Oak Ridge, Tennessee.
- Lavis, R. S., and Chalmers, B., 1958, "On Zone Refining of Metals and the Growth of Alloy Metal Single Crystals," Presented at A.E.C. Conference on Solid State Metallurgy, University of Pennsylvania.
- Faris, J. P., 1958, *Appl. Spectroscopy*, **12**, 6.
- Feher, G., 1957, *Bell System Technical Journal*, **36**, 449-484.
- Hamaguchi, H., Reed G. W., and Turkevich, A., 1957, *Geochimica and Cosmochimica Acta*, **12**, 4, 337-347.
- Holt, B. D., 1959, *Anal. Chem.*, **31**, 1, 51-54.
- Inghram, M., 1959, *Methods of Experimental Physics*, Vol. 6, Part A, Sec. 2.2.3., Academic Press.
- Laitinen, H. A., 1959, Private Communication.
- Leddicotte, G. W., et al., 1959, "The Use of Neutron Activation Analysis in Analytical Chemistry," Paper No. 927, Second United Nations International Conference on the Peaceful Uses of Atomic Energy, Geneva, Switzerland.
- Liebhavsky, H. A., and Winslow, E. H., 1958, *Anal. Chem.*, **30**, 580.
- Mainke, W., 1959, Private Communication.
- Nikelly, J. G., and Cooke, W. D., 1957, *Anal. Chem.*, **29**, 933.
- Rosie, D. J., and Cooke, W. D., 1955, *Anal. Chem.*, **27**, 1360.
- Strain, H. H., 1958, *Anal. Chem.*, **30**, 4, 620.

REVIEWS

1. *Basic Chemistry in Nuclear Energy*, **23**, Proceedings of the Second United Nations International Conference on the Peaceful Uses of Atomic Energy, 1953.